



Machine learning methods for predicting the admissions and hospitalisations in the emergency department of a civil and military hospital

Hugo Álvarez-Chaves¹ · Pablo Muñoz¹ · María D. R-Moreno^{1,2}

Received: 17 February 2023 / Revised: 18 April 2023 / Accepted: 18 April 2023 /

Published online: 4 July 2023

© The Author(s) 2023

Abstract

Hospitals' Emergency Departments (ED) have a great relevance in the health of the population. Properly managing the ED department requires to optimise the service, while maintaining a high quality care. This trade-off implies to properly arrange the schedule for the personnel, so the service can duly attend all patients. In this regard, a key point is to know in advance how many patients will arrive to the service and the number that should be derived to hospitalisation. To provide such information, we present the results of applying different algorithms for forecasting ED admissions and hospitalisations for both seven days and four months ahead. To do this, we have employed the ED admissions and inpatients series from a Spanish civil and military hospital. The ED admissions have been aggregated on a daily basis and on the official workers' shifts, while the hospitalisations series have been considered daily. Over that data we employ two algorithms types: time series (AR, H-W, SARIMA and Prophet) and feature matrix (LR, EN, XGBoost and GLM). In addition, we create all possible ensembles among the models in order to find the best forecasting method. The findings of our study demonstrate that the ensembles can be beneficial in obtaining the best possible model.

Keywords Emergency department · ED forecasting · ED management · Machine learning · Military · Civil hospital

1 Introduction

The Emergency Department (ED) is a strategic service in hospitals and plays a critical role in patients' healthcare. In countries like Spain, where the health care is a public service, patients arrive to the ED without appointment. This poses a challenge for managing the human force and resources of the hospital, so the service provides quality care without cost

✉ Hugo Álvarez-Chaves
hugo.alvarezc@uah.es

¹ Universidad de Alcalá, EPS, ISG, Madrid, Spain

² The Hague, The Netherlands

overrun, when, at the same time, the population has a good perception of the health care system. There are two main aspects to evaluate the ED service: quality and efficiency. On the one hand, quality refers to the time each patient spends in the ED. The patients' waiting time is a sensible factor because of its relationship with potentially fatal consequences. The assistance pressure is related to the mortal rate of the patients Cowan and Trzeciak (2004); Richardson (2006); Guttman et al. (2011), and hence the service management should prevent the collapse of the department by reducing the number of readmissions Mardini and Raš (2019); Raš (2002) or adjusting the resources of the service. On the other hand, efficiency involves the economical cost of the service operation. Therefore, the hospital's administration tries to optimally deal with resources; otherwise, the hospital's viability could be adversely affected.

Service managers should plan the resources required based on the patients' attendance at any specific time. However, forecasting the influx of patients usually depends on the experience and knowledge of the staff. This knowledge may be biased and contradictory among different workers and hospitals, which may introduce errors in the planning process. Furthermore, this planning should be considered in advance to arrange the hospital personnel's schedule. Usually, managers work with two planning horizons: long-term and short-term. The long-term planning establishes a baseline of required resources months in advance that could later be modified in the short term (daily or weekly).

In this work, we aim to model the flow of patients arriving at the ED and inpatients from the service. For this purpose, we will use three years of anonymised data from the emergency department of the *Hospital Central de la Defensa Gómez Ulla* (HCDGU) located in Madrid, Spain. Our dataset contains a total of 207,416 patient records from January 2015 to December 2017. To model the patients flow we use the most common approaches in the literature (Gul and Celik, 2020; Wargon et al., 2009), i.e. time-series and feature-matrix, and the most appropriate time frame for the ED planning. In our case, the selected time frame will be the official shifts to manage the service workers' schedules (3 shifts of 8 hours each day: morning, afternoon and night). Also, patients derived to hospitalisation are handled daily (inpatients require resources from other services, which is beyond the scope of this paper), so this aggregation is used as well. Thus, our study applies two aggregations to the admissions series: shifts of the workers and daily. It is relevant to mention that our work presents the results from a special context, such as a civil and military hospital, but also provides the estimated number of patients for each work shift, which is not common in the literature but very useful for the ED planning.

In order to satisfy the planning constraints for our hospital, we will implement two different forecasting horizons. The long-term method will follow the standard hospital planning requirement, which is four months in advance, while the short-term method would be used to make corrections to long-term predictions, considering seven days ahead. In both methods, we use different Machine Learning (ML) and statistical models such as Autoregressive (AR) (Hyndman and Athansopoulos, 2021), exponential Holt-Winters (H-W) (Hyndman and Athansopoulos, 2021), Seasonal AutoRegressive Integrated Moving Average (SARIMA) (Hyndman and Athansopoulos, 2021), Prophet (Taylor and Letham, 2018), Linear Regression (LR) (Yan and Su, 2009), ElasticNet (EN) (Zou and Hastie, 2005), eXtreme Gradient Boosting (XGBoost) (Chen et al., 2017), Generalized Linear Models (GLM) (McCullagh and Nelder, 2019), and ensembles among them.

The paper has been organised in the following way: Section 2 provides an overview of the related literature. Section 3 details the dataset for the study and its preprocessing in order to adapt to our methodology, which is explained in Section 4. The experimental results are

presented in Section 5 and discussed in Section 6. Finally, Section 7 outlines the conclusions extracted from this work.

2 Literature review

The use of data for the ED service forecast is a popular research topic (Wargon et al., 2009; Gul and Celik, 2020). In recent years, most relevant studies are focused on ML methods following different approaches based on time series or feature matrix.

The time-series approach has been highly considered in the ED forecast. In 1988, Milner (1998) implements an ARIMA method for modelling the ED admissions. Since then, many studies used ARIMA or its variants (such as ARMA, SARMA, SARIMA, or SARIMAX). More recently, works are focused on Exponential Smoothing, AR models or Prophet (Boyle et al., 2012; Hertzum, 2017; Whitt and Zhang, 2019; Gul and Celik, 2020; Choudhury and Urena, 2020; Rocha and Rodrigues, 2021; Park, 2019). In addition to these classical techniques, Deep Learning (DL) has emerged in order to solve the ED forecasting problem. Rocha and Rodrigues (2021) use Long Short-Term Memory Neural Networks (LSTM), while Sharafat and Bayati (2021) created the network called PatientFlowNet in order to forecast the patients' flow in the ED services, including arrivals, treatment and discharges.

The feature-matrix approach appeared in order to adapt some ML methods, such as GLMs (Wargon et al., 2010; Xu et al., 2016) or XGBoost (Rocha and Rodrigues, 2021; Petsis, 2022), to the time series data from the ED. Rocha and Rodrigues (2021) decomposed some characteristics of the admissions series such as year, month, day of the week, or some previous admissions values so they can add exogenous variables to the patients' time series (McCarthy et al., 2008; Wargon et al., 2010; Marcilio et al., 2013; Ekström et al., 2015; Álvarez-Chaves et al., 2021). In this way, McCarthy et al. (2008) concluded that some calendar variables could be useful for modelling. Wargon et al. (2010) applied GLMs with also some calendar variables such as holidays or days of the week. Marcilio et al. (2013) employed weather conditions in addition to calendar variables. Ekström et al. (2015) made use of the region medical website visits in order to forecast ED attendance for the following day. These exogenous variables affect in different ways among distinct hospitals due to the localisation. Other studies used Gradient Boosted Machines (GBM), logistic regressions or decision trees (Poole et al., 2016; Graham et al., 2018) but over historical records for each patient, which in some countries is infeasible due to privacy regulations.

Other studies try to merge both approaches. Xu et al. (2016) presented a hybrid model that sequentially combines ARIMA with LR or Artificial Neural Networks (ANN) to improve the accuracy of a single forecasting model for two and a half years of data. Yucsan et al. (2020) applied a similar methodology over another dataset with only one year of data, concluding that ARIMA-ANN is the most accurate model. Therefore, the combination of different approaches and models may outperform the accuracy of a single algorithm.

The ED forecasting studies often cover other topics, being one of them the inpatients forecasting (Gul and Celik, 2020). Abraham et al. (2009) concluded that it is possible to forecast the occupancy of the wards in the short term (a week ahead), and Peck et al. (2012) implemented a methodology using the information from triage and the number of patients hospitalised from the ED, using GLM with a logit link function and comparing it with expert knowledge. He et al. (2021) set the focus on the inpatients' Length Of Stay (LOS) and developed an ANN-based Multi-task Learning model (ANNML) to predict the inpatient flow and LOS. Bertsimas et al. (2021) made use of interpretable ML methods (regressions

and tree-based) to forecast the LOS of the patients and the discharge destination, including hospitalisation.

Privacy restrictions make data collection a challenge. A large number of works in this field use aggregated records by time on the scale of days, weeks, months, or years (Gul and Celik, 2020). However, the number of studies with a fine-grained scale is reduced. Boyle et al. (2012) took five years of data from two Australian hospitals for training and more than twenty hospitals for testing and found a lower ED attendance in regional areas than urban areas. Similarly, Hertzum (2017) employed three years of data from four Danish hospitals and Rocha and Rodrigues (2021) used ten years of a Portuguese ED service achieving the best performance results on an hourly scale. An objective and reliable comparison among different studies would be difficult to perform because of the distinct data used and the differences in people's habits from different geographical locations.

Considering the forecasting for different time horizons, we can cite the work of Jilani et al. (2019). They implemented a short-term forecasting (four weeks ahead) by making daily aggregations and splitting the series by days of the week, and a long-term predictor (four months ahead) by using monthly aggregations. Boyle et al. (2012) or Rocha and Rodrigues (2021) employed similar aggregations (four or eight hours) but providing only long-term predictions.

This work differs from existing studies on several points. First, we focus on both the admissions and inpatients series. Second, we use the official shifts of the service workers for aggregations for the admissions series. Third, time-series and feature-matrix approaches are used in our work and combined to generate hybrid models (ensembles), finding the best possible model for both series, allowing for an accurate evaluation of the performance of all the models used in the same context. Fourth, this study performs long and short term forecasts based on hospital requirements with the same time aggregations. Finally, our dataset comes from a civil and military hospital, providing a new context to the field.

3 Data description

The historical records of the ED is the starting point for any patient influx forecasting. Our study uses data from the HCDGU located in Madrid, Spain. It has the peculiarity of being a military hospital, but some departments also admit civilian patients, such as the ED. Therefore, we are dealing with a context where civilians arrive as in a public hospital but it retains the military designation and organization.

Our dataset contains three years of admissions and hospitalisations, from January 1, 2015, until December 30, 2017. In that period, the number of patients attended by the ED service was 207,416, being each admission a record in our dataset. In the following subsections we detail how we adapt the data for its use on time-series and feature-matrix techniques. Also, Section 3.3 provides an insight of the distributions when splitting the dataset by shifts.

3.1 Time series

Our dataset has time-series format so preprocessing was not necessary. We rely on three different time series based on the time aggregation used: two for the admissions and one for the inpatients. For the patients admissions we consider daily aggregations, i.e., the number of all the patients' arrivals on one specific day, and the official shifts aggregations. The official shifts of the service staff are: from 8:00 a.m. to 3:00 p.m., known as the first shift or

morning shift; from 3:00 p.m. to 10:00 p.m., named second shift or afternoon shift; and from 10:00 p.m. to 08:00 a.m., called third shift or night shift. Regarding the inpatients series, we only consider daily aggregations as our goal is to help the ED managers in their planning (hospitalizations are handled in a different way by other services).

Performing an additive decomposition of each series into trend, seasonality and residuals, allows us to observe a strong seasonality pattern by weeks and days. Figure 1 shows the additive decomposition for admissions series aggregated by days, where an upward trend is observed, but which decreases during certain periods of the year, such as in the summer. We decided to keep the raw series in order to maintain the original data due to the absence of pronounced outliers. The weekly seasonality can be better appreciated in Fig. 2, where the seasonal component is presented for the admissions and inpatients series using daily aggregations. The seasonality component for inpatients series has less variability, with about four-patient difference between high and low peaks hence these series seems less stable than admissions series. In addition to the weekly seasonality, we also perceive a remarkable daily seasonality for admissions series aggregated by shifts. The autocorrelation plot displaying the daily seasonality can be seen in Fig. 3 (bottom), where there is a strong correlation between the same shift on previous days. It should be noted that the series also have a yearly seasonality because we observe that registers decrease in the summer and increase in the last quarter of the year. Nevertheless, this seasonality could not be proved because our dataset only contains three years of data. We will employ the seasonality information extracted from our exploratory analysis for those time-series algorithms which implement seasonal components.

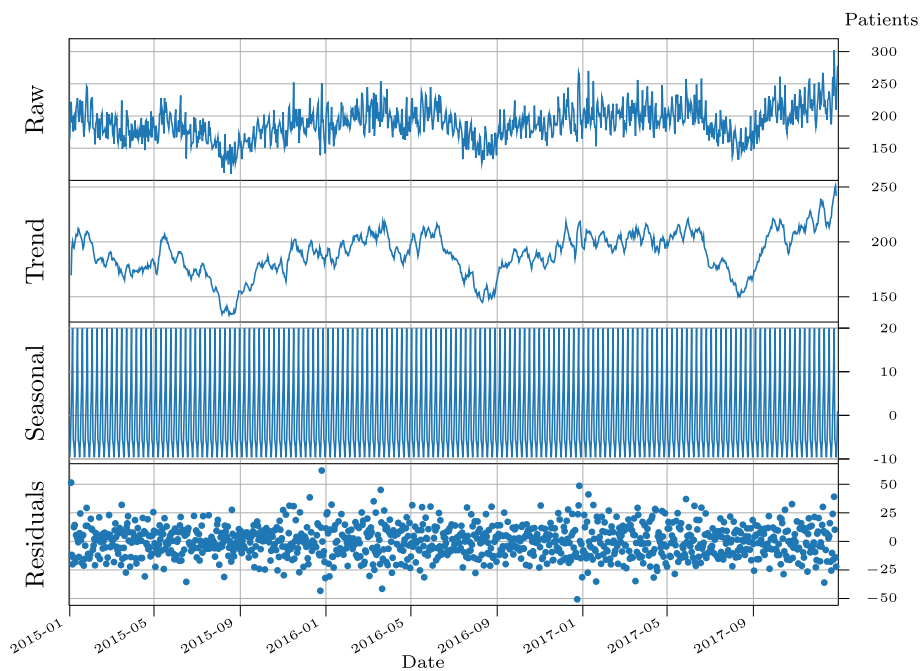


Fig. 1 Additive decomposition for the admissions series aggregated by days. Top plot the raw patients' admissions series and (from top to bottom) its additive decomposition in trend, seasonality and residuals components. The trend component could indicate a yearly seasonality which decreases in during the summer and increases at the end of the year. The seasonal pattern shows a strong weekly seasonality

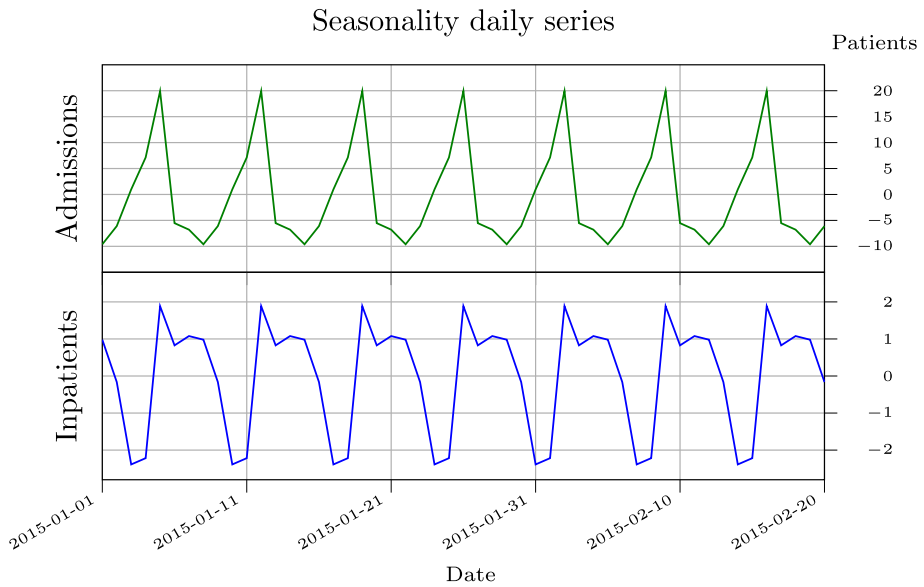


Fig. 2 Seasonal component for admissions and inpatients series, both in daily aggregations. The plot shows a weekly pattern that will be used to modelling the time series

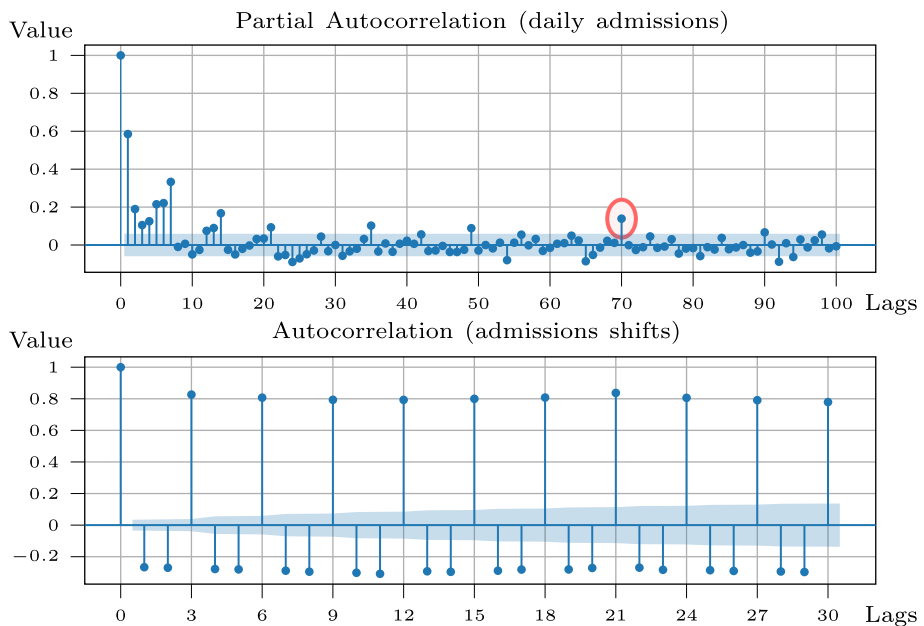


Fig. 3 Partial correlation for the admissions series in daily aggregations (top). There is a significant correlation until the lag 70 (selected value for the feature-matrix approach). Autocorrelation plot for the admission series in shift aggregations (bottom). There is a strong correlation between the same shift on previous days

3.2 Feature matrix

Feature-matrix approaches employ the same data described in the previous section, but these algorithms cannot extract the time-series relations by themselves. Therefore, the data conversion is necessary to maintain the temporal relationship. For our purpose, we apply a transformation similar to the one performed in Álvarez-Chaves et al. (2021).

The dataset transformation process uses the previous time series values as the description feature for the current value. Thus, the target variable is represented by the time series value at a concrete instant in time for this approach. Then, we obtain the feature by shifting the values of the time series, thus the target variable at a specific point in time is used as a feature variable for the following time values. Therefore, the feature matrix consists of the current value and a defined number of shifting movements of the previous records in the time series. The number of movements selected was 70 days since we have found a significant correlation until this lag (see Fig. 3). For applying this process we need to remove the first 70 rows due to the undefined values in their previous values, being inconsistent with most of the applied algorithms, as can be seen in Fig. 4.

3.3 Distributions

One of the objectives of this study is to obtain the best possible forecasting method among those used. Having this in mind, we analyse the data distributions. Splitting the series by shifts, we can appreciate some differences among shifts distributions. Morning and afternoon shifts present a stable behaviour in the admissions series with a wide range of values in the distributions. In contrast, the night shift values are around a lower range with two peaks in the distribution. Figure 5 shows the Probability Density Function (PDF) by using Kernel Density Estimation (KDE) with Gaussian kernels and the Scott's Rule (Scott, 2010) for each shift with data normalised to 0-1 range.

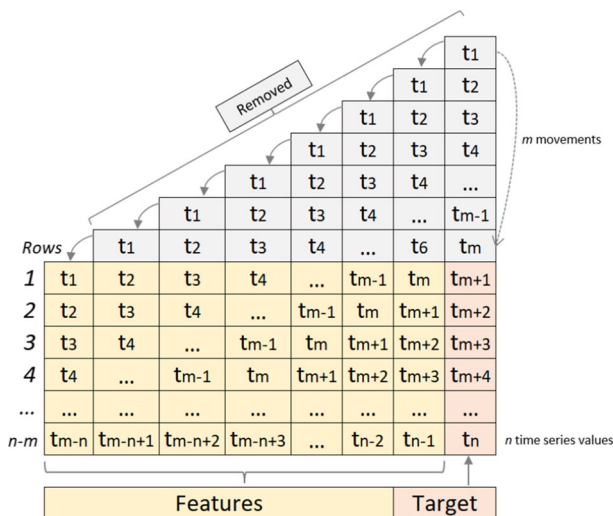


Fig. 4 Time-series to feature-matrix dataset conversion process. We select 70 days for the movements. Therefore the first 70 rows need to be removed

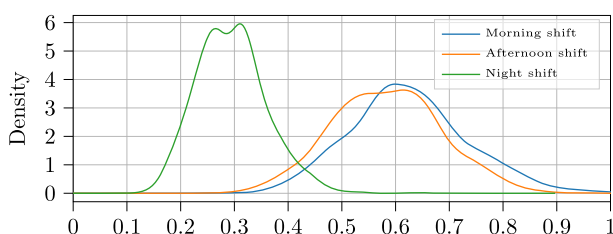


Fig. 5 PDF of the admission series divided by shifts and normalized to 0-1. It is observed the morning shift concentrates the highest patient influx and the highest peak (reason why higher values close to 1 are found). In contrast, the night shift shows the minimum mean of number of patients and has two peaks in its distribution

This finding may be interesting to be used in our implementation. In addition to the entire series, we use the division of the admissions series by shifts to see if we can improve forecasting accuracy.

4 Methodology

For forecasting the patients influx to the ED department we use time-series and feature-matrix approaches. On the one side, the time-series algorithms make use of the time-series data extracting time relationships among values. On the other side, the feature-matrix data is used for classical ML algorithms. Both approaches require different steps, so we divide our methodology into a time-series approach and a feature-matrix approach. While the time-series approach requires few steps to get the best possible results, the one using classical ML algorithms requires more efforts to fit the algorithms to improve the accuracy. The algorithms employed in our study are AR, exponential H-W, SARIMA, and Prophet for the time-series approach, and LR, EN, XGBoost, and GLM for the feature-matrix approach. With all models fitted, the last step is to create ensembles in order to try to improve each model's single results.

Although they are two different approaches, both aim to make long-term and short-term predictions. To this end, we split our dataset into 86% of the time series for training (from January 1, 2015, to August 7, 2017) and the other 14% for testing (from August 8, 2017, to December 30, 2017). We make this partition in order to meet the requirements of the hospital for long-term planning (more than four months) by forecasting the entire test period. In addition, some algorithms require a hyperparameter tuning, such as SARIMA, Prophet, EN, XGBoost, and GLM. In those cases, we use a subset from April 8, 2017 to August 7, 2017 of the training set for validation. To select the best hyperparameters we used the hyperopt optimization (Bergstra et al., 2013) with TPE algorithm (Bergstra et al., 2011) for a maximum of 1.000 iterations.

For short-term predictions, the hospital requires weekly forecasts. In this case, we make use of a process known as walk forward (Kaastra and Boyd, 1996) in order to simulate the performance of the model when new data arrives. Our walk forward implementation uses an extensible window (instead of a sliding window), so the training set grows in each iteration of the process. We train the model each step to forecast the following week and, when the process ends, compute the metrics for all the test set. Algorithm 1 depicts our *walk forward* implementation.

In the following subsections we detail the time series and features-matrix approaches used, concluding the section with the creation of the ensembles.

Algorithm 1 Pseudocode of the walk forward algorithm implemented in the evaluation process of short-term modelling.

```

period ← 7 days
forecasts ← list()
copytest ← copy(testset)
set model hyperparams
while testset is not empty do
    model.training(trainingset)
    forecast ← model.forecasting(period)
    forecasts.append(forecast)
    trainingset.append(testset(dataperiod))
    testset.remove(dataperiod)
end while
computemetrics(copytest, forecasts)

```

4.1 Time-series approach

Time-series methods use the raw historical records and are able to take advantage of the time relationship among values. In this approach, we apply the more common models in ED time series modelling (Gul and Celik, 2020). The selected algorithms are: AR, exponential H-W, SARIMA, and Prophet. Furthermore, a *Naïve* algorithm is implemented to set a baseline for our study.

4.1.1 Naïve

The *Naïve* modelling is based on the weekly seasonality of the time series. Thus, the latest weekly values are used to forecast the following week for short term predictions. Therefore, we replicated every week in order to get the next week's forecasts, so every forecast value is similar to the same time one week before. As our study also faces long-term forecasting, the baseline model employs the yearly seasonality detected. In this way, we replicate the value from the same day of the week of the previous year, corresponding to 364 days before.

4.1.2 Autoregressive and exponential Holt-Winters

The first time-series model employed is AR (Hyndman and Athansopoulos, 2021). The AR algorithm uses consecutive values to model the time series. This algorithm finds the best linear regressor by using previous consecutive values to forecast the next value.

Exponential models are also popular in time-series modelling. Among these types of models, we selected Holt-Winters (H-W) (Hyndman and Athansopoulos, 2021) due to the characteristics of our dataset. H-W is able to fit the seasonality and the trend of a time series, which is our case. This algorithm can be used in two different forms; additive or multiplicative. In our case, we selected the additive form because it obtains the best accuracy in the experiments.

4.1.3 SARIMA

The third time-series algorithm applied is SARIMA (Hyndman and Athansopoulos, 2021). This algorithm combines AR and Mean Average (MA) models and allows us to exploit

cyclical patterns detected due to its seasonal part. For all series, we use the weekly seasonality for the algorithm.

Prior to utilising SARIMA models, it is necessary to establish its components. In our work, we tried the `auto.arima` algorithm (Hyndman and Khandakar, 2008) by using the step-wise process to select the best model, but it could not find good models (it falls in local minima, stopping the process early). Therefore, the models were selected by testing a large number of combinations until we achieve an acceptable accuracy. Among all the possible SARIMA models for each time series, we selected SARIMA(6,1,1)(3,0,5)[7] for the admissions series using daily aggregations, and SARIMA(18,0,5)(2,0,1)[21] for the admissions series using shifts aggregations. Splitting admissions series by shifts, the chosen models were SARIMA(6,1,1)(3,0,5)[7] for morning and afternoon shifts, and SARIMA(6,1,1)(3,0,3)[7] for the night shift. At last, the model selected for the inpatient series in daily aggregations was SARIMA(6,1,1)(1,0,2)[7].

SARIMA models can also handle exogenous data. In this way, we introduced some description features of the date to the model, such as the day of the week, day of the month, month of the year, week of the year, day of the year, days in the month, and week of the month.

4.1.4 Prophet

Prophet (Taylor and Letham, 2018) is the last time-series algorithm selected in our study. Prophet decomposes the time series into trend, seasonality and holidays or special events. In our study, the holidays component is not used in order to enable the comparison with the other algorithms. Moreover, the Prophet library does not contain the Spanish holidays. Nevertheless, we added the same description features of the date to the model as in SARIMA.

In the case of Prophet algorithm, we have to set the hyperparameters before its application. The hyperparameters space are presented in Table 1. We decided this search space because of the impact on the modelling task and the algorithm author recommendations. In addition, a daily, weekly and yearly seasonality is set to the algorithm.

4.2 Feature-matrix approach

Feature-matrix enables to model our problem selecting only those previous values with the most substantial influence for each series. Therefore, we implemented a procedure, based on forward feature selection, that only chooses the variables that provide useful information for those algorithms that need it. We use three classical regression models: Linear Regression (LR), ElasticNet (EN) and Generalized Linear Model (GLM). We also use a tree-based regressive model, eXtreme Gradient Boosting (XGBoost).

Table 1 Hyperopt space for Prophet algorithm

Hyperparameters	Range of values
Hyperparameters	Range of values
changepoint prior scale	[0.001 - 0.5]
changepoint range	[0.8-0.95]
seasonality mode	additive or multiplicative
number of changepoints	[25-30]
seasonality prior scale	[0.001-0.1]

The feature-matrix models selected can handle exogenous data. For this reason, we also introduce the same date description features as in SARIMA and Prophet (day of the week, day of the month, month of the year, week of the year, day of the year, days in the month, and week of the month).

In the following subsections, we explain the steps required to adapt the algorithms for the feature selection task.

4.2.1 Linear regression

The LR (Yan and Su, 2009) model is not able to select features by itself. For this reason, we define an iterative process similar to the one performed in Álvarez-Chaves et al. (2021) and depicted in Algorithm 2. We use a threshold of 0.001 and the R^2 score to evaluate and decide the addition of variables. After using the process, no further steps are required.

Algorithm 2 Feature-matrix linear regression approach pseudocode (Álvarez-Chaves et al., 2021).

```

unselected_features ← list of all features
selected_features ← list()
threshold ← 0.001
selected_features.append(previous_time_series_value)
baseline ← 10fold_CV(selected_features, train)
while unselected_features is not empty do
    impact_list ← list()
    for each unselected_feature in unselected_features do
        test_feature_list ← list(selected_features, unselected_feature)
        value ← 10fold_CV(test_feature_list, train)
        impact_list.append(value)
    end for
    best_feature ← best_value(impact_list)
    if best_feature > (baseline + threshold) then
        selected_features.append(best_feature)
        baseline ← 10fold_CV(selected_features, train)
    else
        break
    end if
end while
fit_model(selected_features, train)

```

4.2.2 ElasticNet

The EN regression model (Zou and Hastie, 2005) can detect unhelpful variables by setting its model parameters to zero. The EN model has three different hyperparameters and the search space for the hyperopt optimization is presented in Table 2. We decided this search space because of the impact on the modelling task based on our experience.

4.2.3 XGBoost

The XGBoost model (Chen et al., 2017) is able to not select useless variables by itself in its construction process. The algorithm builds trees by selecting the most relevant variables

Table 2 Hyperopt space for EN algorithm

Hyperparameters	Range of values
alpha	[0.0001-100]
l1 ratio	[0.1-1.0]
number of iterations	[0-1000]

in order to make the best forecasts. XGBoost has numerous hyperparameters; the chosen search space is presented in Table 3. We chose these hyperparameters due to the impact on the performance of the model in our experience and for our specific problem. Furthermore, we decided to set gbtree (instead of dart) as the booster employed as it gets the better performance in our tests.

4.2.4 GLM

The GLM models (McCullagh and Nelder, 2019) generalise the linear models and fit better to the independent variables whose distribution differs from the normal. They use a function over the regression, called link function. The link function is the hyperparameter for this algorithm, thus we search it among logit, probit, cauchy, log, cloglog, power, inverse power, identity, nbinom, sqrt, and inverse squared by testing all of them and selecting the one with the best accuracy. When the search of the hyperparameter is performed, then we use the same process as for LR and described in Algorithm 2.

4.3 Ensembles

The models mentioned above have different characteristics in terms of data fitting. Ensembling different algorithms to create a more complex model may improve the predictions of every single model. Initially, the ensemble approach was to combine the models with the lowest correlation in order to balance their performance. However, after conducting multiple experiments and verifying that the computational cost was small, we opted to combine all models to test all possible combinations. This approach allows us to not only test the models with the lowest correlation, but also those that may have a non-linear or monotonic relationships. Thus, the methodology implemented to make these ensembles consists of combining each one of the models with the rest of them, using subsets from two to all the models. Each

Table 3 Hyperopt space for XGBoost algorithm

Hyperparameters	Range of values
eta	[0.01-1.0]
gamma	[0-9]
max depth	[3-18]
subsample	[0.5-1.0]
number of estimators	[10-100]
columns sample by tree	[0.3-1.0]
min child weight	[1-10]
reg alpha	[40-180]
reg lambda	[0.0-1.0]

ensemble averages the forecasts of its component models, giving the same weight to the forecasts of all of them.

Additionally to the ensembles by models, our dataset allows to make a different type of ensemble. We focus the new ensemble on the admission series of shifts taken independently, i.e. creating a new series for each shift as we detailed in Section 3.3 (morning, afternoon, and night series). Then, we fit every single model, and also perform all the ensembles to get the best forecasting method for each shift series. Finally, we merge the predictions in order to recreate the forecasts for the original shift series. This aims to obtain the best model considering the differences in the shift distributions.

5 Experimental results

This section presents the results of applying the different algorithms to our ED admission dataset to forecast the patients influx in two temporal horizons. In this regard, to ease the assessment, we provide each forecast (short and long-term) in a dedicate section. The results are discussed in Section 6.

The metrics for the models' evaluation are the most used in the literature: MAPE, MAE and R^2 . In addition, we compute the **Pearson correlation** (ρ) in order to check the linear correlation between the forecasts and real data.

5.1 Long-term forecast

Table 4 provides the metrics for forecasting the test set for the long-term predictions. Only the best models of each approach (time series, feature-matrix and ensembles) and the baseline

Table 4 Long-term metrics for the best models for time series, features-matrix and ensembles

Series	Aggregation	Model	MAPE	MAE	R^2	ρ
Admissions	Daily	Baseline	11.26	23.84	0.05	0.67
		Prophet	7.96	17.17	0.48	0.85
		EN	11.28	24.37	0.08	0.82
		Ensemble	7.96	17.17	0.48	0.85
	Shifts	Baseline	16.33	10.94	0.64	0.84
		Prophet	12.40	8.17	0.80	0.91
		LR	15.16	10.63	0.67	0.91
		Ensemble	12.40	8.17	0.80	0.91
	IS	Ensemble	11.54	7.52	0.82	0.92
Inpatients	Daily	Baseline	23.95	4.37	-0.14	0.38
		Prophet	20.74	4.07	-0.01	0.43
		EN	20.46	3.83	0.12	0.46
		Ensemble	20.16	3.82	0.11	0.47

IS aggregation corresponds to the ensemble model by shifts. All Pearson correlations are significant ($p_{value} < 0.05$). Highlighted in bold the best result for each aggregation

model for all the series and aggregations analysed are presented. We can see that each approach overcomes the baseline results for almost all cases.

Considering the admissions series in daily aggregations, our baseline gets a value of R^2 near zero, although the *Pearson correlation* shows a positive correlation of 0.67. The best time-series method is Prophet; its results improves significantly the baseline, getting an MAE of 17.17, MAPE of 7.96, R^2 of 0.48 and a high ρ value of 0.85. This model is able to fit better to the trend and seasonality of the data than the others algorithms. The combination of Prophet and ARIMA through ensembles resulted in the best performance for this approach, but it did not surpass the results obtained by the Prophet model alone, decaying to 20.11 of MAE, 9.27 of MAPE, 0.32 of R^2 , and 0.84 of ρ . In this case, the time-series approach has the better accuracy.

Regarding the admissions series in shifts aggregations, the comparison between approaches is similar to the daily aggregations. The best results are also obtained by the Prophet model, being selected as the only model for the ensemble again. Prophet obtains 12.40, 8.17, 0.80 and 0.91 for the MAPE, MAE, R^2 and ρ respectively. In this case, the best feature-matrix algorithm is the LR, slightly improving the metrics of the baseline, but obtaining a ρ value similar to the Prophet model. In this case, the ensemble consists of Prophet and LR models, but its metrics do not succeed in enhancing the Prophet model, even though they maintain a similar ρ value. Nevertheless, the ensemble created by using different models for each shift improves the results with a MAPE of 11.54, MAE of 7.52, R^2 of 0.82, and ρ of 0.92. The ensemble in this case is composed by the Prophet model for each single shift. The addition of other models to this ensemble does not improve the forecasts.

Our algorithms obtains better predictions for the inpatients series than for the admissions series; this can be seen in all metrics with the exception of the MAE, which is better since the inpatients series has lower values. The baseline presents the characteristic of getting a value of -0.14 for R^2 and 0.38 for ρ . Analysing our baseline forecasts, they follow the yearly trend of the series (which is consistent with the model definition and the exploratory analysis), but the weekly pattern does not seem to be as pronounced as in the admissions series. All the models improve the baseline, being the EN model the best single model for this case with metric values of 20.46, 3.83, 0.12, and 0.46 for MAPE, MAE, R^2 and ρ respectively. The Prophet model is still the best time-series model, but in this case the R^2 decays until -0.01 value. The ensemble model for the inpatients is composed by the AR, LR and EN models, slightly improving the results for the single models.

5.2 Short-term forecast

We computed the metrics for the short-term forecast using the walk forward algorithm presented in Section 4. Table 5 shows the best metrics for each series, aggregation and approach. In this case, all baselines are outperformed by both approaches, being the best results obtained by the time-series.

Regarding the admissions series aggregated by days, time-series models improve from an MAE of 17.17 in the long-term to 14.53 in the short-term, remaining the Prophet model as the best among those that follow this approach. Nevertheless, in this case the Ensemble model is composed of averaging SARIMA and Prophet, achieving the best result with an MAE of 13.73 and R^2 of 0.67 (ρ remains almost constant). For the feature-matrix models, EN highly outperforms its results in comparison to the long-term approach improving from an MAE value of 24.37, and R^2 of 0.08 to 14.70, and 0.65 respectively.

Table 5 Short-term metrics for the best models for time series, features-matrix and ensembles

Series	Aggregation	Model	MAPE	MAE	R^2	ρ
Admissions	Daily	Baseline	9.04	18.33	0.47	0.73
		Prophet	6.84	14.53	0.63	0.85
		EN	7.10	14.70	0.65	0.82
		Ensemble	6.58	13.73	0.67	0.84
	Shifts	Baseline	15.08	9.63	0.72	0.86
		SARIMA	11.60	7.61	0.82	0.92
		GLM	12.26	7.78	0.82	0.91
		Ensemble	11.39	7.36	0.83	0.92
	IS	Ensemble	11.23	7.25	0.84	0.92
	Inpatients	Baseline	27.68	5.01	-0.54	0.21
		H-W	20.01	3.74	0.17	0.46
		GLM	20.47	3.77	0.17	0.46
		Ensemble	20.03	3.70	0.18	0.46

IS aggregation corresponds to the ensemble model by shifts. All Pearson correlations are significant ($p_{value} < 0.05$). Highlighted in bold the best result for each aggregation

For the admissions series in shift aggregations, the ensemble of H-W, SARIMA, Prophet and GLM obtains the best predictions with a MAPE of 11.39, MAE of 7.36 and R^2 of 0.83. Nevertheless, the accuracy among all approaches are very similar. Focusing on each approach, SARIMA is the best model for time-series approach and GLM for the feature-matrix approach. In addition, a relevant result is obtained in the case of the combination of a model for each single shift. In this case, the ensemble slightly improves the best result for the entire shift series model, achieving MAPE of 11.23, MAE of 7.25, R^2 of 0.84 and ρ of 0.92. This ensemble is composed by the average of Prophet and LR for the morning shift; H-W, SARIMA, Prophet, and LR for the afternoon shift; and Prophet and EN for the night shift.

Concerning the inpatients series, the baseline obtains worse results for short-term compared to the long-term: with R^2 from -0.14 to -0.54 and MAE from 4.37 to 5.01. In contrast, the rest of the models improve their results. We get the H-W as the best single model with a MAPE of 20.01, MAE of 3.74, R^2 of 0.17 and ρ of 0.46. In this case, the feature-matrix model selected is GLM, obtaining similar results to H-W. The best ensemble is obtained by merging AR, H-W, and GLM models, achieving a MAPE of 20.03, MAE of 3.70, R^2 of 0.18 and ρ of 0.46.

6 Discussion

Considering the results presented in the previous section, we will discuss how we can use the proposed algorithms' predictions to manage our reference hospital.

Modelling the hospitalisation data is a challenging task, as it reflects the stochastic nature of the arrival of the most critical patients. These patients require urgent care and immediate

attention, unlike those who seek the service due to the unavailability of primary care appointments. This is consistent with the exploratory analysis presented above that shows a low degree of seasonality in this series. Therefore, the models derived from hospitalisations would have limited applicability for the service management. Nevertheless, we obtain better predictions for the admissions series, likely due to strong patterns observed in the exploratory analysis. The prediction algorithms have a direct utility to assist the service managers to better understand the patients influx, so they can arrange the ED hospital resources more than four months in advance.

We can exploit the good predictions obtained by the admission series aggregated by shifts to enhance the long term service personnel schedule. Considering that the holidays calendar is known in advance, ED managers can use such forecast to easily allocate the personnel in the most demanded shifts, so it also improve work-life balance of the workers. In this regard, the forecasts adapt to the intraday patterns of the service, which are the most defined, as shown in Fig. 6.

Using the two time horizons for the admissions series in daily aggregation is specially relevant as the forecast achieves the better accuracy. As shown in Fig. 7, the corrections made by the short-term forecasts facilitate a better adaptation to the trend of the series. This enables the managers to adjust the initial long-term forecasts. For the series of shifts, we achieve similar results for both long and short term, with a slight improvement, which indicates that the algorithms used might be more susceptible to noise despite they better capture the trend.

Technically, the outcomes of the long-term baselines confirm the presence of an yearly seasonal pattern in the admissions series, confirming our intuition. Another interesting aspect is that the feature matrix methods are outperformed by the time series methods in all cases except for the long-term inpatients series. The main reason behind this could be the weak seasonality of the inpatients series, causing that the time series models could not take advantage of its main characteristic, i.e., time patterns. For this case, a model that can select its variables, and use the date description, finds better patterns to fit. Conversely, in the short-term horizon the time series models obtain similar results overcoming the difficulties they face for the long term.

We also notice that the best overall metrics are achieved by the ensemble aggregated by shifts. Despite sacrificing the continuity of the series by modelling each shift independently, it improves with respect to each single model. This may have its foundation in the difference

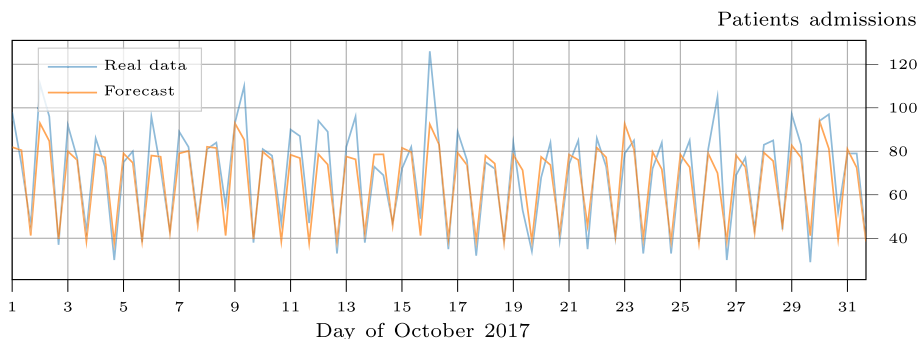


Fig. 6 October forecasts of the ensemble model by independent shifts for the admission series in the long term (best overall model). The mean is shown to compare the forecasting matching to the uptrend of the series

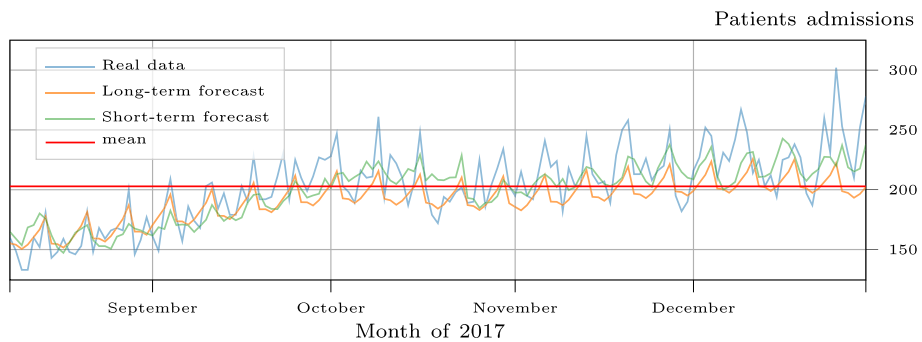


Fig. 7 Forecasts obtained by the best models for the admissions series in a daily aggregation (Prophet in the long term and ensemble in the short term). The mean is shown to compare the forecasting matching to the uptrend of the series

of the shifts distributions discussed in the exploratory analysis. Modelling each distribution independently could simplify how the general model captures the internal patterns of the shifts.

It is significant to mention that the emergency attendance depends on the health system and the habits of the people in a particular society. These factors are reflected in the patterns that we have seen in our series, mostly in the admissions one. Furthermore, we have employed data that come from a civil and military hospital, so we are dealing with a rather special case.

After discussing with ED service experts about how to deploy the models for supporting the decision making, we are developing an application that shows two planning scenarios. The first one would display the evolution of patient influx for the long term (four months in advance), where the service managers could assign the fixed resources they consider necessary for that period (nurses per shift and hospital equipment). The second one provides short-term planning (seven days in advance), where the managers can reallocate non-fixed staff and consumable resources to avoid reducing the quality of the service in unexpected situations. To facilitate the usability of the models without requiring expert knowledge from the managers, the best way is to deploy a user-friendly web application. It displays the different predictions and sends alerts to notify of a sudden change in the patient forecast, indicating to the manager the recommendation to re-evaluate the situation.

7 Conclusions

Our study describes an analysis of the ED forecasting for a Spanish civil and military hospital by using different algorithms based on time series and feature matrix approaches, and creating ensembles among them. In order to optimise the applicability of the study for the hospital managers, we have analysed the series of patients who were admitted to the ED (admissions series), and the patients who were transferred to the hospital wards after arriving to the ED (inpatients series). With the first series, we have achieved satisfactory predictions, which can be used to improve the ED's resources allocation on a daily basis, as well as to better schedule the personnel's shifts. In contrast, the models applied to the inpatients series did not attain such satisfactory forecasts due to the reduced seasonality of the series. Nevertheless, the low mean error in the inpatients forecasting should be considered for further studies considering the patients referral between departments.

Through modelling the two time horizons, we have noticed that applying the long-term corrections using the short-term forecasts improves the prediction, so the ED managers can better react to unexpected high demand. However, with long-term planning, the tasks such as procuring material or setting the schedules for the permanent staff would be feasible, while requiring auxiliary staff when the corrected forecast indicates so. This also improves work-life balance of the workers.

Considering that we only have 3 years of historical data (for both training and testing), we have applied the feasible methods having in mind this issue. We aimed to exploit the daily, weekly and yearly seasonality that we discerned through the exploratory analysis. For this task, the best individual algorithm was Prophet. Nevertheless, the combination of different models by generating ensembles can lead to better outcomes in certain cases. While not all the cases may see improvement, the creation of ensembles presents an intriguing avenue for achieving the best possible model.

The conclusions drawn in this study are contingent on our data set and its origin, in this case the HCDGU, Madrid, Spain. The outcomes of the modelling task and its applicability relies heavily on this context. In spite of this statement, the methodology employed, as well as the identification of the features of our context, can provide a foundation for other studies or analogous applications with other data.

Acknowledgements The authors would like to express their gratitude to the clinical support of Dr. Helena Hernández Martínez, Dr. María Isabel Pascual Benito and Mr. Francisco López Martínez. The authors also want to acknowledge the support of the Hospital Central de la Defensa Gómez Ulla, Madrid, Spain

Author Contributions All authors equally contributed to this paper.

Funding This study was funded by Ministerio de Ciencia e Innovación (PID2019-109891RB-I00). Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability The data set used cannot be published due to data protection restrictions.

Declarations

Competing interests The authors have no competing interests or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Ethical Approval The use of the data from this study has been approved by the Ethical Committee of the Hospital Central de la Defensa Gómez Ulla, Madrid, Spain. Data were anonymised prior to use.

Limitations In our analysis, we rely on the limitations of the data source. On the one hand, we only have data from a single hospital with particular characteristics, i.e., the hospital provides medical care to military and civilian population in Madrid, a large city with many hospitals. On the other hand, our dataset contains only three years of ED records, so the presence of larger seasonalities, e.g., yearly, cannot be properly evaluated in our algorithms.

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

References

- Abraham, G., Byrnes, G. B., & Bain, C. A. (2009). Short-term forecasting of emergency inpatient flow. *IEEE Transactions on Information Technology in Biomedicine*, 13(3), 380–388. <https://doi.org/10.1109/TITB.2009.2014565>
- Álvarez-Chaves, H., Barrero, D.F., Cobos, M., et al. (2021). Patients forecasting in emergency services by using machine learning and exogenous variables. In: *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 167–180. <https://doi.org/10.1007/978-3-030-91100-315>
- Bergstra, J., Bardenet, R., Bengio, Y., et al. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems* 24. Available at <https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf>
- Bergstra, J., Yamins, D., Cox, D. (2013). Making a Science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: *Proceedings of the 30th International Conference on Machine Learning*, 115–123. Available at <https://proceedings.mlr.press/v28/bergstra13.html>
- Bertsimas, D., Pauphilet, J., Stevens, J., et al. (2021). Predicting inpatient flow at a major hospital using interpretable analytics. *Manufacturing & Service Operations Management*. <https://doi.org/10.1287/msom.2021.0971>
- Boyle, J., Jessup, M., Crilly, J., et al. (2012). Predicting emergency department admissions. *Emergency Medicine Journal*, 29(5), 358–365. <https://doi.org/10.1136/emj.2010.103531>
- Chen, T., He, T., Benesty, M., et al. (2017). xgboost: eXtreme Gradient Boosting. R package version 06-4 1(4), 1–4. Available at <https://cran.microsoft.com/snapshot/2017-12-11/web/packages/xgboost/vignettes/xgboost.pdf>
- Choudhury, A., Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. *British Journal of Healthcare Management* 26(1), 34–43. <https://doi.org/10.12968/bjhc.2019.0067>
- Cowan, R. M., & Trzeciak, S. (2004). Clinical review: Emergency department overcrowding and the potential impact on the critically ill. *Critical Care*, 9(3), 1–5. <https://doi.org/10.1186/cc2981>
- Ekström, A., Kurland, L., Farrokhnia, N., et al. (2015). Forecasting emergency department visits using internet data. *Annals of Emergency Medicine*, 65(4), 436–442. <https://doi.org/10.1016/j.annemergmed.2014.10.008>
- Graham, B., Bond, R., Quinn, M., et al. (2018). Using Data Mining to Predict Hospital Admissions From the Emergency Department. *IEEE Access* 6(10), 458–10,469. <https://doi.org/10.1109/ACCESS.2018.2808843>
- Gul, M., & Celik, E. (2020). An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Systems*, 9(4), 263–284. <https://doi.org/10.1080/20476965.2018.1547348>
- Guttmann, A., Schull, M. J., Vermeulen, M. J., et al. (2011). Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario. *Canada. Bmj*, 342, d2983. <https://doi.org/10.1136/bmj.d2983>
- He, L., Madathil, S. C., Servis, G., et al. (2021). Neural network-based multi-task learning for inpatient flow classification and length of stay prediction. *Applied Soft Computing*, 108(107), 483. <https://doi.org/10.1016/j.asoc.2021.107483>
- Hertzum, M. (2017). Forecasting hourly patient visits in the emergency department to counteract crowding. *The Ergonomics Open Journal* 10(1). <https://doi.org/10.2174/1875934301710010001>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. Melbourne: OTexts. , ISBN: 9812834109
- Hyndman, R.J., Khandakar, Y. (2008). Automatic Time Series Forecasting: *The forecast Package for R*. *Journal of Statistical Software* 27, 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Jilani, T., Housley, G., Figueredo, G., et al. (2019). Short and long term predictions of hospital emergency department attendances. *International Journal of Medical Informatics*, 129, 167–174. <https://doi.org/10.1016/j.ijmedinf.2019.05.011>
- Kaasra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215–236. [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9)
- Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Academic Emergency Medicine*, 20(8), 769–777. <https://doi.org/10.1111/acem.12182>
- Mardini, M. T., & Raš, Z. W. (2019). Extraction of actionable knowledge to reduce hospital readmissions through patients personalization. *Information Sciences*, 485, 1–17. <https://doi.org/10.1016/j.ins.2019.02.006>

- McCarthy, M. L., Zeger, S. L., Ding, R., et al. (2008). The Challenge of Predicting Demand for Emergency Department Services. *Academic Emergency Medicine*, 15(4), 337–346. <https://doi.org/10.1111/j.1553-2712.2008.00083.x>
- McCullagh, P., & Nelder, J. A. (2019). Generalized linear models. *Routledge, New York*. <https://doi.org/10.1201/9780203753736>
- Milner, P. (1988). Forecasting the demand on accident and emergency departments in health districts in the trent region. *Statistics in Medicine*, 7(10), 1061–1072. <https://doi.org/10.1002/sim.4780071007>
- Park, J., Chang, B., & Mok, N. (2019). 144 Time series analysis and forecasting daily emergency department visits utilizing facebook’s prophet method. *Annals of Emergency Medicine*, 74(4), S57. <https://doi.org/10.1016/j.annemergmed.2019.08.149>
- Peck, J. S., Bennenay, J. C., Nightingale, D. J., et al. (2012). Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine*, 19(9), E1045–E1054. <https://doi.org/10.1111/j.1553-2712.2012.01435.x>
- Petsis, S., Karamanou, A., Kalampokis, E., et al. (2022). Forecasting and explaining emergency department visits in a public hospital. *Journal of Intelligent Information Systems*, 59(2), 479–500. <https://doi.org/10.1007/s10844-022-00716-6>
- Poole, S., Grannis, S., Shah, N.H. (2016). Predicting emergency department visits. *AMIA Summits on Translational Science Proceedings 2016*, 438. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5001776>
- Raš, Z. (2022). Reduction of hospital readmissions. *Advances in Clinical and Experimental Medicine* 31(1), 5–8. <https://doi.org/10.17219/acem/144413>
- Richardson, D. B. (2006). Increase in patient mortality at 10 days associated with emergency department overcrowding. *Medical Journal of Australia*, 184(5), 213–216. <https://doi.org/10.5694/j.1326-5377.2006.tb00204.x>
- Rocha, C. N., & Rodrigues, F. (2021). Forecasting emergency department admissions. *Journal of Intelligent Information Systems*, 56, 1–20. <https://doi.org/10.1007/s10844-021-00638-9>
- Scott, D. W. (2010). Scott’s rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 497–502. <https://doi.org/10.1002/wics.103>
- Sharafat, A.R., Bayati, M. (2021). PatientFlowNet: A deep learning approach to patient flow prediction in emergency departments. *IEEE Access* 9(4)5, 552–45,561. <https://doi.org/10.1109/ACCESS.2021.3066164>
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- Wargon, M., Guidet, B., Hoang, T., et al. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal*, 26(6), 395–399. <https://doi.org/10.1136/emj.2008.062380>
- Wargon, M., Casalino, E., & Guidet, B. (2010). From model to forecasting: A multicenter study in emergency departments. *Academic Emergency Medicine*, 17(9), 970–978. <https://doi.org/10.1111/j.1553-2712.2010.00847.x>
- Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. *Operations Research for Health Care*, 21, 1–18. <https://doi.org/10.1016/j.orhc.2019.01.002>
- Xu, Q., Tsui, K. L., Jiang, W., et al. (2016). A hybrid approach for forecasting patient visits in emergency department. *Quality and Reliability Engineering International*, 32(8), 2751–2759. <https://doi.org/10.1002/qre.2095>
- Yan, X., & Su, X. (2009). *Linear regression analysis: Theory and computing*. Singapore: World Scientific. , ISBN:9812834109
- Yucesan, M., Gul, M., & Celik, E. (2020). A multi-method patient arrival forecasting outline for hospital emergency departments. *International Journal of Healthcare Management*, 13(sup1), 283–295. <https://doi.org/10.1080/20479700.2018.1531608>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>