Time Series Models for Predicting the Number of Patients Attending the Emergency Department in a Local Hospital

Silvia Aguirre, Jon Kerexeta and Moises Espejo

July 1, 2024

Time series models for predicting the number of patients attending the emergency department in a local hospital

Silvia Aguirre[1], Jon Kerexeta[1,2], Moises Espejo[3]

[1]Vicomtech Foundation, Basque Research and Technology Alliance(BRTA), Donostia, 20009, Spain.
[2]Biodonostia Health Research Institute, San Sebastián, 20014, Spain.
[3]Asunción Klinika, Tolosa, 20400, Spain.

Contributing authors: silvia.agirre11@gmail.com; jkerexeta@vicomtech.org; mespejo@clinicadelaasuncion.com;

Abstract

The daily influx of patients at emergency departments (EDs) is highly unpredictable and a major cause of overcrowding in hospitals. This study aims to provide decision-making support and establish a shared situational awareness among medical and administrative personnel. By accurately forecasting daily attendances, this project attempts to effectively reduce overcrowding issues and improve overall patient care. To address this issue, this study focuses on studying different models to predict the number of visits to the emergency departments and investigating the factors affecting daily demand. Hospitals can benefit from accurately forecasting the number of patients arriving at the ED, allowing for early planning and mitigating overcrowding. As the subject of the study, a real database collected from Asunción Klinika from 2004 to 2022 was examined in Tolosa, Gipuzkoa. For this purpose, models such as ARIMA, LSTM and GRU are proposed. The study revealed that weekly patterns as well as calendar and meteorological information have an impact on the volume of daily patient arrivals. Over the years, several forecasting models using time series analysis have been proposed to address this challenge. Results showed that hybrid models outperformed the others in terms of the Mean Absolut Error metric (MAE). Predictions have yielded an average daily error of 5.2 individuals, which accounts for 13%.

Keywords: Time series forecasting, ED demands, prediction models, RNN-LSTM,RNN-GRU, daily patient arrivals, ARIMA

Introduction

The demand for medical care has increased rapidly over the last decades, becoming crucial the management of medical staff and patient flows. EDs represent the core mission of healthcare facilities. They are considered the main gateway to the hospital, as they are an almost mandatory step for all patients before admission to most hospital services. With the high demand for emergency services, saturation can occur and thus reduce the quality of medical services. In this regard, forecasting daily emergency attendances is vital to mitigate problems of overcrowding and help ED managers allocate available resources effectively.

Despite the efforts made in recent years to define an emergency care model, the emergency department remains a top priority challenge where improvement opportunities exist. At present, there exists a lack of information regarding the estimation of the number of patient arrivals to ED. The current model generates waiting queues and a single flow of care, which can limit service capacity and negatively impact patients' perception of the quality of service due to long waiting times and possible errors in resource optimization.

This study aims to provide decision-making support and establish a shared situational awareness among medical and administrative personnel. By accurately forecasting daily attendances, the project aims to effectively mitigate overcrowding issues and improve overall patient care.

Related work

There is an extended literature on forecasting ED patient arrival models using time series analysis methods. The arrival of patients is influenced by multiple variables and is highly unpredictable. The conventional linear model fails to capture the nonlinear relationship among these variables. However, a nonlinear approach can be developed to achieve more accurate results in a complex dynamic system [1]. As Vidra K. Sudarshan [2] correctly mentions, there are several models in particular which have been extensively explored for the ED forecasting development. Such models are: Linear Regression, ARIMA, Seasonal ARIMA and Exponential Smoothing (ES)[3],[4]. However, ARIMA is the preferred method for analyzing time series due to its robust mathematical background [5],[6].

According to recent research papers [7],[8], Machine Learning (ML) and Deep Neural Network (DNN) based techniques have shown to perform best when applied to solve the time series prediction. Deep learning models can better capture the dependencies when data is not directly amenable to feature engineering. Among DNN techniques, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are the most used DNN models for time series predictions as they use sequence input data and can learn from a large temporal window.

Existing literature on forecasting ED arrivals suggests the inclusion of calendar variables such as time of day, day of the week, and month as predictors [9] [10],[11]. Some studies [12],[2] propose that maximal, minimum, mean, temperature and humidity are the environmental variables that can influence the ED arrivals. Others [2],[13] also use calendar indicators such as national holidays and school holidays. However, there have been times when the use of precipitation as a variable may have introduced uncertainty to the model and almost no improvement to the accuracy of the forecast [2],[1]. These investigations indicate that visits to the ED exhibit seasonal patterns, with calendar variables playing a significant role [5],[14].

This study proposes the implementation of different models to enhance the findings studied in the state of the art. Firstly, the ARIMA model, which has undergone extensive research and possesses a robust mathematical foundation. Additionally, RNN models, specifically LSTM and GRU, as they are considered the most advanced models and have demonstrated remarkable performance in recent years. To achieve this, environmental and calendar variables will be incorporated into the analysis, examining the impact on a weekly, monthly, and annual basis. Moreover, hybrid models will be explored, which combine the strengths of both ARIMA and RNN models to find the best possible outcome.

Materials and Methods

Database

The data to be analyzed consisted of patient data who attended the emergency department from 2004 to 2022. Analyzing the data, it has been observed in Figure 1 that there is an increasing trend of 0.75 every year, as well as a weekly and yearly seasonality.

As it can be seen in Figures 2 the monthly peak demand is observed in January with 43 arrivals per day, compared to 35.6 arrivals per day in August. On Mondays, (see Figure 3) the arrivals per day reach up to 44.35, while on Thursdays it decreases to 38.7, on average. Furthermore, it has also been studied the arrival time of patients during the day; there is one peak for increased demand which is at around 11am.

A wide range of daily patient arrivals was noticed. By excluding the COVID-19 period it was observed that the majority of the data lied within the range of 30-56. As basic descriptive analysis of the data, the range of incomes of the database was between 8-72. While the mean income per day was 40.42 . In order to ensure more accurate predictions, outliers were substituted with the number of arrivals from the previous day to ensure the preservation of the seasonality and peaks within the dataset (See Figure 4).

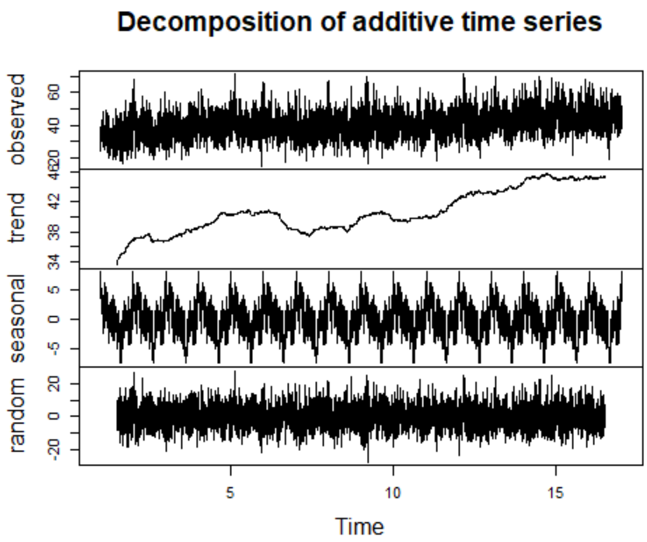

Fig. 1 Description of the data.

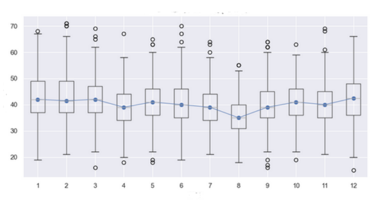

Fig. 2 Monthly patient arrivals

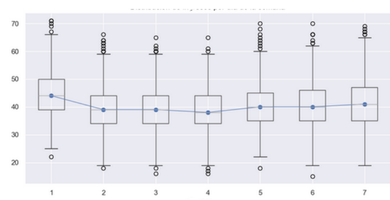

Fig. 3 Weekly patient arrivals

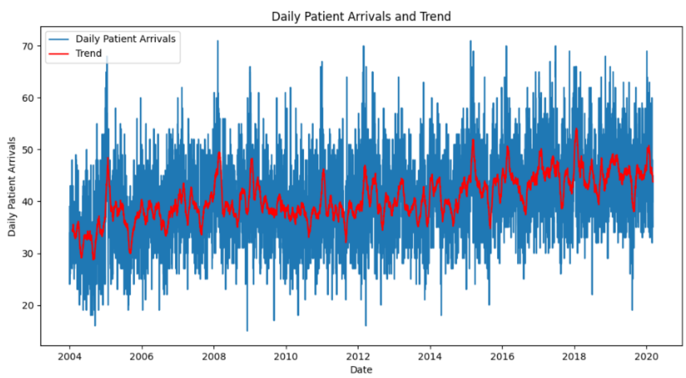

Fig. 4 Daily patient arrivals.

In addition, environmental data was extracted from the Open-Meteo (Open-source weather API) [15]. The data was extracted from Tolosa, Gipuzkoa (Latitude 43.13 and Longitude -2.08) from the 'historical weather API' section.

Regarding weather data, the obtained variables were the following: maximum temperature(°C), minimum temperature(°C), mean temperature(°C), shortwave radiation(MJ/m^2), wind speed(km/h), rain(mm), relative humidity(%), pressure mean(Pa) and snowfall(mm). Moreover, several national and local festivals were collected in a boolean variable.

The final database contained a total of 27 variables and 6861 rows.

Methods

Time series models such as ARIMA typically estimate patient visits by considering three factors: long-term trends, cyclical changes (such as season, weather, or day of the week), and the impact of unexpected random events. Due to their ease of use, implementation, and interpretation in modeling and forecasting future variables, time series methods are widely applied in various domains.

The ARIMA model provides an approach to time series and forecasting that is one the most widely used methodologies to time series analysis. This model takes into account the existing dependence between the data, meaning that each observation at a given moment is modeled based on previous values. Furthermore, it can incorporate a cyclic or seasonal component. Its generic name ARIMA(p,d,q), derives from its three components AR (AutoRegressive), I (Integrated), and MA (Moving Averages). Calculations and analysis are used to determine the order parameters (p, d, q) for fitting an ARIMA model [16]. The mathematical expression for this model is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - ... - \theta_q \epsilon_{t-q} \quad (1)$$

where X_t represents the variable value at time t, and ϕ and θ are the model parameters for the autoregressive and moving average terms, respectively. The residual term ϵ_t represents random disturbances that cannot be predicted, and it is generally assumed that the error terms ϵ_t are independent.

Morever, ARIMAX models are an extension of ARIMA models that incorporate exogenous variables to better explain the behavior of the target variable, which in this case is the arrival of patients.

Among different models which represent the data equally well the simplest one is preferred, i.e. the model with the fewest parameters. The figure below presents a flowchart of the Box Jenkins model previously introduced [17] , [18].

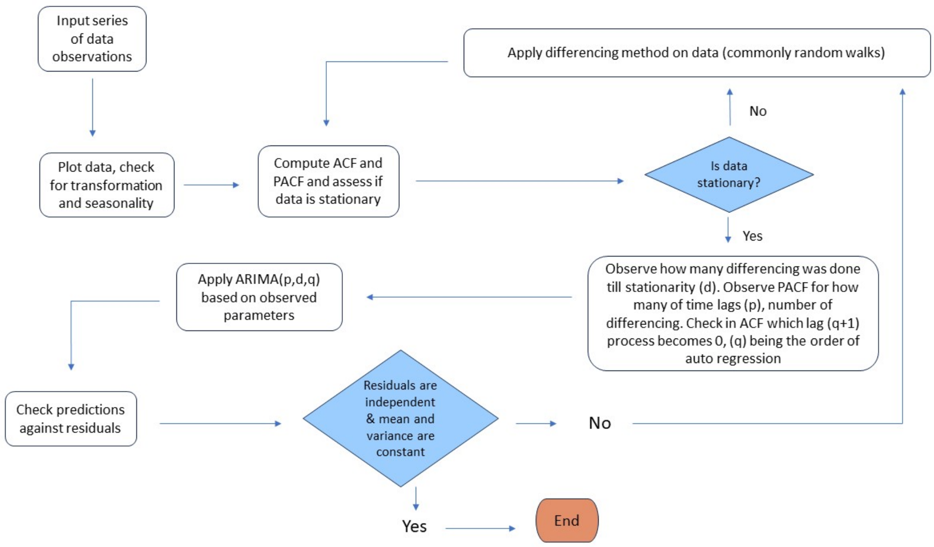

Fig. 5 Box Jenkins fit iterative fitting model [17] , [18]

Gated Recurrent Neural Networks

Gated recurrent neural network models have been introduced in the literature as a way to effectively capture dependencies in time series data with large time step distances. LSTM and GRU are the two primary types of RNNs that are capable of retaining information over extended periods of time [19] , [20].

Long Short Term Memory (LSTM)

Long Short Term Memory is one of the most popular and efficient RNN that is capable of learning long term dependencies. It was first proposed by Hochreiter and Schmidhuber in 1997 [21] to address the vanishing gradient problem [22]. This model is composed of cells, with three main gates known as the input gate, forget gate, and output gate. These gates are responsible for regulating the flow of information into and out of the memory cell, the core of the LSTM.
The basic procedure of the LSTM network at time step t can be mathematically described as follows:

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{2}$$

First, the forget gate (F_t) is designed to decide what information should be discarded from the previous cell state C_{t-1} . Secondly, the new information should be conditionally stored in the cell state. A tanh layer is selected to create the new memory, while the input gate determines what should be added to the cell state.

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{3}$$

$$\widetilde{C_t} = tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \tag{4}$$

Third, the cell state of the current time step C_t will be updated by combining the candidate memory C_t with the long term memory C_{t-1}.

$$C_t = F_t * C_{t-1} + I_t * \widetilde{C_t} \tag{5}$$

In conclusion, the output gate is formulated to determine which data will be converted from the cell state into the current hidden layer data.

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{6}$$

$$H_t = O_t * tanh(C_t) \tag{7}$$

where W_{xi}, W_{xf}, W_{xo} and $W_{hf},\ W_{hi},\ W_{ho}$ are weight parameters and $b_i,\ b_o,\ b_f$ are bias parameters.

Gated recurrent unit (GRU) model

The Gated Recurrent Unit (GRU) model, proposed by Cho et al [23], is considered a variation of LSTM models. It shares similarities but has fewer parameters. The resulting model is simpler than standard LSTM models, making it more compact and efficient.
The reset gate R_t is computed by

$$R_t = \sigma(X_t W_{xr} + H_{hr} + b_r) \tag{8}$$

This model computes candidate hidden states that aid in the computation of subsequent hidden states. The new remember $\widetilde{H_t}$ is generated by R_t with a $tanh$ layer:

$$\widetilde{H_t} = tanh(X_t W_{xh} + W_{hh}(R_t * H_{t-1}) + b_h) \tag{9}$$

GRU eliminates the remember gate and forget gate present in LSTM and introduces Z_t to replace them. By using the update gate Z_t at the current time step, the previous time step's hidden state $Ht - 1$ can be combined with the candidate hidden state H_t of the current time step to compute Ht for that specific time step t. It is computed by

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \tag{10}$$

Lastly, the hidden state value is updated by

$$H_t = (1 - Z_t) * H_{t-1} + Z_t * \widetilde{H_t} \tag{11}$$

where similar to LSTM, W_{xr}, W_{xr} and W_{hr}, W_{hz} are weight parameters and b_r, b_z are bias parameters.The output is computed through a fully connected layer with a *sigmoid* function serving as its activation function.

Model Evaluation Criteria

In this study, AIC, MAE and RMSE are employed to evaluate the precision of ED attendance forecasting, while MSE is used as a loss function for Deep Learning models. These metrics are mathematically formulated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{12}$$

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

where $\hat{y}_i$ is the forecast value, y_i is the true value, and n is the sample number. Intuitively, if the MAE is 5, it means that there is an average margin error of 5 people.

$$AIC = -2\,ln\,L(\widehat{\beta}, \widehat{\sigma}^2)\frac{2(p+q+1)n}{(n-p-q-2)} \tag{14}$$

where $\widehat{\beta}$ and $\widehat{\sigma}^2$ denote the sample parameters of the given ARMA process and L is the maximized value of the likelihood function for the model. Intuitively, the term $\frac{2(p+q+1)n}{(n-p-q-2)}$ can be considered a penalty term to discourage over-parameterization.

Results

The ARIMA model that demonstrated the best fit was ARIMA(2,0,4)(0,1,1)[7]. To ensure a good fit for the model, several tests were conducted. Initially, the augmented Dickey-Fuller test of stationarity yielded a P-value ($P > 0.01$), indicating that the null hypothesis of the existence of a unit root was rejected. This suggests that the actual series is stationary and hence, that any desired ARIMA model can be applied. The Box–Ljung test on residuals resulted in a P-value of 0.238 ($P > 0.01$), indicating a normal distribution and independence of the residuals.
Best results were obtained when incorporating weather variables such as maximum, minimum, and mean temperature, average rainfall, wind speed, daily relative humidity and holidays. Introducing exogenous variables has led to a MAE of 5.49 per day.

In order to build RNN models, data was split into three subsets: %80 as train set, %10 as validation and %10 as test set.
The LSTM and GRU candidate models were chosen and validated based on the evaluation metrics discussed in the preceding section 3, namely MAE and RMSE.
Numerous combinations of hyperparameters were tested through a grid search to achieve optimal results. Furthermore, variables such as weekdays, months, mean average of the recent month, and arrivals of the last year were also included.

Table 1 summarizes the hyperparameters employed in the modeling and fitting of both models. The table 2 provides a summary of the comparison between the LSTM

Hyperparameters	**LSTM**	**GRU**
Neuron units	[64/64/32/24]	[64/64/32/24]
Dropout	[0.1/0.05/0.05]	[0.2/0.2/0.1]
Activation function	ReLu, Linear	ReLu, Linear
Optimizer	Adam	Adam
Loss function	MSE	MSE
Epochs	150	150
Batch size	16	16
Number of layers	4 (2 dense layes)	4 (2 dense layers)
Learning rate	$4 \cdot 10^{-4}$	$4 \cdot 10^{-4}$
Look back	30	30

Table 1 Hyperparamters and learning algorithms.

Metrics	LSTM	GRU	SARIMA	SARIMAX
RMSE	7.27	7.64	6.98	6.94
MAE	5.38	5.4	5.53	5.49
Loss function(MSE)	0.33	0.37	-	-
Total parameters	286184	35720	7	13

Table 2 Comparison results for daily arrivals

and GRU candidate models in terms of daily patient arrivals at the ED. Both models have the same architecture and sample, except for the dropout. Based on the results shown in Table 2, it can be observed that the GRU model performs slightly worse than LSTM.

However, with the aim of improving the result, this study proposes to introduce detrended data as input to the RNN models. This idea arose from the desire to combine the ARIMA and RNN models, which are referred to as hybrid models in theory.

To begin with, ARIMA was applied on the original data. Then, its residuals were introduced as input for the LSTM and GRU models, along with the exogenous variables, aiming to achieve a lower error. Ultimately, it led to the most effective approach up to now. The mean absolute error decreased to 5.2 in GRU and 5.3 in LSTM. Figures 7 and 6 display the predicted and actual values for daily patient arrivals.

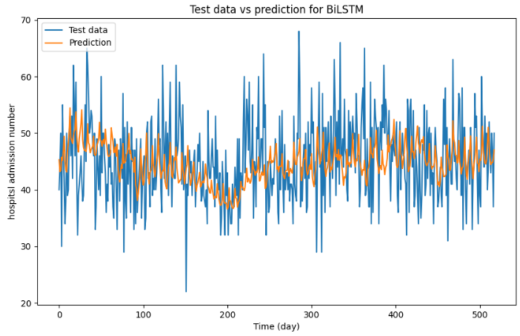

Fig. 6 Predicted and actual daily patient arrival of the LSTM model

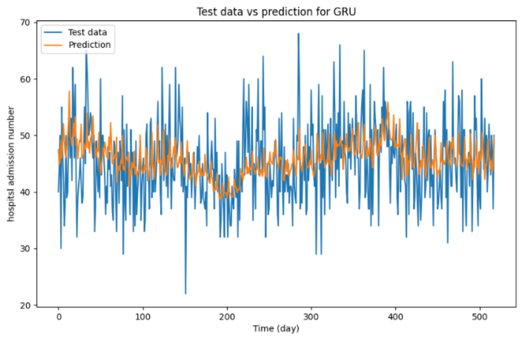

Fig. 7 Predicted and actual daily patient arrival of the GRU model

Discussion

In the beginning of the study, ARIMA was expected to provide the most precise forecasts, based on its success in similar problems in the state-of-the-art. However, the results suggest that the hybrid LSTM and GRU models proposed improved accuracy in short-term forecasting, with a MAE error (see Section 3) above 5%.

Temperature, rainfall and holidays consistently played significant roles. Moreover, weekdays and months do impact in the number of arrivals to the EDs.

The gathered information is highly beneficial for the clinic when determining the appropriate number of medical personnel needed to address the demands of the week, month... This enables them to effectively plan and organize the workforce to accommodate the surge in patient visits and prevent overload.

As regarding future lines, the main objective is to examine the instances when the model experiences the most significant failures. To investigate it further, we have established a technique that will automatically search for news reports on the selected specific days in the Tolosaldea area. This approach will facilitate to assess whether there have been any anomalies on those days or if there exists any specific factor that could be incorporated into the model as a variable.

Conclusions

This study focuses on the development of different models to predict the number of patient visits to the emergency department (ED). Real data from the ED of the hospital of Tolosa (Gipuzkoa) was used to develop the models.
Specifically, ARIMA and two RNN-based deep learning approaches (LSTM and GRU) are presented for predicting patient attendances in EDs.

Based on the obtained results, it can be inferred that both calendar and meteorological data are related with the arrival of ED patients. Nevertheless, further research is needed to determine if the obtained accuracy is sufficient for practical use in a clinical setting.

References

[1] Zhang, Y., Zhang, J., Tao, M., Shu, J., Zhu, D.: Forecasting patient arrivals at emergency department using calendar and meteorological information. Applied Intelligence **52**(10), 11232–11243 (2022) https://doi.org/10.1007/s10489-021-03085-9

[2] Sudarshan, V.K., Brabrand, M., Range, T.M., Wiil, U.K.: Performance evaluation of emergency department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. Comput Biol Med **135** (2021) https://doi.org/10.1016/j.compbiomed.2021.104541

[3] Wargon, M., Guidet, B., Hoang, T.D., Hejblum, G.: A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal **26**(6), 395–399 (2009) https://doi.org/10.1136/emj.2008.062380

[4] Harrou, F., Kadri, F., Sun, Y., Khadraoui, S.: Monitoring patient flow in a hospital emergency department: Arma-based nonparametric glrt scheme. Health Informatics J **27**(2) (2021) https://doi.org/10.1177/14604582211021649

[5] Choudhury, A., Urena, E.: Forecasting hourly emergency department arrival using time series analysis. British Journal of Health Care Management **26**(1), 34–43 (2020) https://doi.org/10.12968/bjhc.2019.0067

[6] Jones, S.S., Thomas, A., Evans, R.S., Welch, S.J., Haug, P.J., Snow, G.L.: Forecasting daily patient volumes in the emergency department. Academic Emergency Medicine **15**(2), 159–170 (2008) https://doi.org/10.1111/j.1553-2712.2007.00032.x

[7] Kadri, F., Abdennbi, K.: Rnn-based deep-learning approach to forecasting hospital system demands: application to an emergency department (2020)

[8] Sharafat, A.R., Bayati, M.: Patientflownet: A deep learning approach to patient flow prediction in emergency departments. IEEE Access **9**, 45552–45561 (2021) https://doi.org/10.1109/ACCESS.2021.3066164

[9] Røislien, J., Søvik, S., Eken, T.: Seasonality in trauma admissions – are daylight and weather variables better predictors than general cyclic effects? PLoS One **13**(2) (2018) https://doi.org/10.1371/journal.pone.0192568

[10] Peck, J.S., Benneyan, J.C., Nightingale, D.J., Gaehde, S.A.: Predicting emergency department inpatient admissions to improve same-day patient flow. Academic Emergency Medicine **19**(9), 1045–1054 (2012) https://doi.org/10.1111/j.1553-2712.2012.01435.x

[11] Tandberg, D., Quails, F.C.: Time series forecasts of emergency department patient volume, length of stay

[12] Carvalho-Silva, M., Monteiro, M.T.T., Sá-Soares, F., Dória-Nóbrega, S.: Assessment of forecasting models for patients arrival at emergency department. Oper Res Health Care **18**, 112–118 (2018) https://doi.org/10.1016/j.orhc.2017.05.001

[13] Whitt, W., Zhang, X.: Forecasting arrivals and occupancy levels in an emergency department. Oper Res Health Care **21**, 1–18 (2019) https://doi.org/10.1016/j.orhc.2019.01.002

[14] Yousefi, M., Yousefi, M., Fathi, M., Fogliatto, F.S.: Patient visit forecasting in an emergency department using a deep neural network approach. Kybernetes **49**(9), 2335–2348 (2020) https://doi.org/10.1108/K-10-2018-0520

[15] API. [Online]. Available: https://openmeteo.substack.com/embed

[16] Helfenstein, U.: Box-jenkins modelling in medical research. Stat Methods Med Res **5**(1), 3–22 (1996) https://doi.org/10.1177/096228029600500102

[17] al., E.E.E.: A comparison of univariate and multivariate forecasting models predicting emergency department patient arrivals during the covid-19 pandemic. Healthcare (Switzerland) **10**(6) (2022) https://doi.org/10.3390/healthcare10061120

[18] Duarte, D., Faerman, J.: Comparison of time series prediction of healthcare emergency department indicators with arima and prophet. Academy and Industry Research Collaboration Center (AIRCC), 123–133 (2019) https://doi.org/10.5121/csit.2019.91810

[19] Schmidt, R.M.: Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview. http://arxiv.org/abs/1912.05911

[20] Medsker, L., Jain, D.L.: RECURRENT NEURAL NETWORKS. B. Raton London New York Washington, ??? (2001). Edited by Design and Applications

[21] Hochreiter, S., Schmidhuber, J.U.: Long short-term memory (Unknown)

[22] Grosse, R.: Lecture 15: Exploding and Vanishing Gradients. [Online]. Available: https://commons.wikimedia.org/w/index.php?curid=321973

[23] Chen, J., Jing, H., Chang, Y., Liu, Q.: Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliab Eng Syst Saf **185**, 372–382 (2019) https://doi.org/10.1016/j.ress.2019.01.006