Expert Systems With Applications 240 (2024) 122496

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Evaluating the impact of exogenous variables for patients forecasting in an Emergency Department using Attention Neural Networks

Hugo Álvarez-Chaves *, Iván Maseda-Zurdo, Pablo Muñoz, María D. R-Moreno

Universidad de Alcalá, Departamento de Automática, Alcalá de Henares, 28805, Madrid, Spain

ARTICLE INFO

Keywords:
Emergency Department
Admissions forecast
Exogenous variables
Deep Neural Network
Attention mechanism
Google Trends

ABSTRACT

Emergency Department overcrowding is a well-known problem. The consequences are long waiting times for patients, reduced service quality, and the potential for increased mortality rates. Moreover, this problem is escalating with the aging of the population and the chronification of diseases. Being aware of such concerns, Emergency Departments started to conduct patient accountability years ago, and now they are ready to take advantage of the gathered data to improve the quality of service to citizens. However, to achieve this goal, they need the technology that enables the generation of predictive models from such data to improve the planning of their resources. In this paper, we face the task of forecasting patient admissions by using Deep Neural Networks, particularly a modern Attention-based model. As well, historical records do not provide a full picture of how patients attend the emergency service. Relying on the possibilities given by the attention mechanism, we also propose the use of exogenous information such as calendar data, weather, air quality, allergens, and information extracted from the web via Google Trends to enhance admissions prediction accuracy. In this regard, we have tested different configurations of the exogenous variables to isolate which ones provide relevant information that improves the model. In our experiments, we have seen that the calendar, weather, and air quality provide the most valuable information, meanwhile, allergens and Google Trends data appear to be hindering the models' performance.

1. Introduction

Overcrowding Emergency Departments (EDs) is a concerning global problem that has become clear during the COVID-19 pandemic. Overcrowding can be defined as the situation in which ED function is impeded primarily because of the excessive number of patients awaiting review, undergoing assessment and treatment, or waiting for departure compared to the capacity of the ED. The consequences are long waiting times for patients, reduced service quality, and the possibility of increases in mortality rates. Several studies revealed that mortality is directly related to the delays in diagnosis and its treatment (Cowan & Trzeciak, 2004; Guttmann, Schull, Vermeulen, & Stukel, 2011; Seymour et al., 2017). As well, ED personnel's work-life balance is a key point in providing a quality service (Healy & Tyrrell, 2011; Suokas & Lönnqvist, 1989); having a working schedule in advance can ensure that professionals have enough family and resting time so they can duly attend to their duties during working hours. Therefore, it is important that the ED service is well managed, properly sizing the service personnel to avoid crowding, delays, and staff burnout, and at the same time not incurring cost overruns.

To achieve this, it is common to rely on management knowledge, considering the historical records to predict the number of patients that will arrive at the service. However, only considering such records could not provide accurate predictions. In this regard, some studies point out that assessing other data can improve the patient's forecast (Diehl, Morris, & Mannis, 1981; Duwalage, Burkett, White, Wong, & Thompson, 2020; Petsis, Karamanou, Kalampokis, & Tarabanis, 2022). For instance, weather data or information about air quality or allergen concentrations can be used, as well as holiday information or data extracted from social networks or Internet service providers. Also, different forecasting techniques have been used, from autoregressive models to Deep Neural Networks (DNN), being the last recently introduced in this domain. An extensive review in this direction has been written by Jiang, Liu, and Ding (2022), briefly discussed in the next section.

In this study, we will focus on improving the forecasting of patient arrivals to the ED of a Spanish civil-military hospital located in Madrid. Particularly, we will employ DNNs with Attention Mechanisms aided by exogenous variables to provide accurate patient arrivals forecast.

* Corresponding author.
E-mail addresses: hugo.alvarezc@uah.es (H. Álvarez-Chaves), ivan.maseda@uah.es (I. Maseda-Zurdo), pablo.munoz@uah.es (P. Muñoz), malola.rmoreno@uah.es (M.D. R-Moreno).

https://doi.org/10.1016/j.eswa.2023.122496
Received 5 July 2023; Received in revised form 16 October 2023; Accepted 5 November 2023
Available online 14 November 2023

In particular, we will use calendar data, meteorological and air quality records, allergens concentrations, and data collected from Google Trends regarding some search terms (selected by ED workers) in the hospital area. Based on these ideas, we have developed a DNN with attention layers that provide better accuracy than other neural networks used in time-series forecasting. We will also conduct a study on how the proposed exogenous variables affect the forecasting capabilities of the network. This is a key point as Machine Learning (ML) techniques are strongly affected by the input data. It is expected that by providing more information to the DNN, in particular to the attention layers, the model is able to enhance the predictions that cannot be explained by the hospital records.

As we will demonstrate, such information will not necessarily improve the forecasts. While modern DNN with Attention Mechanisms can focus on relevant information, there are limitations to the extent of information that they can assess. Pushing the limits may lead to injecting noise instead of useful data and thus, reducing the accuracy of the forecast. To properly assess this issue and determine which information can be used to enhance the service quality we have evaluated combinations of different variables (calendar info, allergens, weather, air quality, and subsets of Google Trends searches) using two different DNNs: a state of the art Long Short Term Memory (LSTM) based on the model proposed by Sudarshan, Brabrand, Range, and Wiil (2021) and our proposed DNN with Attention layers. Assessing the predictions outcome we see that the attention mechanism clearly outperforms LSTM-based networks. As well, calendar information is the most relevant exogenous variable, while allergens and meteorological records provide some benefits as well. However, Google Trends information reduces the accuracy as the information is not easy to correlate with ED admissions, even when we restrict the search terms used to the most relevant ones based on experimental evaluation.

This paper is structured as follows: the next section presents related works. Section 3 provides an insight into the data used: ED records, weather, allergens, and calendar data used, and Google Trends information. As well, it presents the LSTM model and our attention DNN. Section 4 provides the assessment of the different models evaluated and a discussion on how our proposal can provide support for ED manager's decisions. Finally, some conclusions are outlined.

2. State of the art

The research of patient admissions in ED has experienced notable growth, with several studies focused on the development of prediction models to estimate the number of patients who will arrive at these services (Gul & Celik, 2020; Jiang et al., 2022; Silva et al., 2022; Sudarshan et al., 2021). In a systematic review conducted by Jiang et al. (2022) several forecasting methods employed for predicting patient demand were collected, identifying that DNN models have gained popularity in recent years. The studies presented used a diverse spectrum of techniques and approaches to model the influx of patients to the ED, encompassing statistical models and Deep Learning techniques. Also, some studies incorporate exogenous variables such as weather, calendar events, and even information retrieved from the Internet. Additionally, the authors emphasize the importance of considering the unique characteristics of each hospital when selecting relevant factors and appropriate modeling methods.

Among the studies presented by Jiang et al. (2022), they follow different classical approaches based on regression analysis and time series methods. On one hand, for the regression methods, Multiple Linear Regression (MLR) was used in some studies such as Ho, To, Koh, and Cheong (2019), Zhang, Zhang, Tao, Shu, and Zhu (2022), Ekström, Kurland, Farrokhnia, Castrén, and Nordberg (2015), and Boyle et al. (2012), among others such as Poisson regression (Marcilio, Hajat, & Gouveia, 2013; McCarthy et al., 2008), Support Vector Regression (SVR) (Yousefi, Yousefi, Fathi, & Fogliatto, 2020; Zhang et al., 2022), or Nonlinear Least Square Regression (Xu, Wong, & Chin, 2013). On the other hand, time series methods were applied by using algorithms such as Autoregressive Integrated Moving Average (ARIMA) (Choudhury & Urena, 2020; Jilani et al., 2019; Rocha & Rodrigues, 2021; Zhang et al., 2022), or different Exponential methods (Boyle et al., 2012; Jones, Thomas, Evans, Welch, Haug, & Snow, 2008; Rema & Sikdar, 2021; Rocha & Rodrigues, 2021). While these models were the first to be employed in addressing the problem and are still being used, it is worth noting that other models have emerged during the last few years. Most of the studies in this field use the Mean Absolute Percentage Error (MAPE) as the evaluation metric for the models. For classical approaches, values ranging from 4.8% to 11.8% are obtained for daily aggregations (Jiang et al., 2022).

Deep Learning techniques have also been employed to model the patient's admissions to the service. Sudarshan et al. (2021) used LSTM and Convolutional Neural Networks (CNN) for this task. Another recent study conducted by Rocha and Rodrigues (2021) used Recurrent Neural Networks (RNN) among other DNN architectures. Zhang et al. (2022) proposed a deep stacked architecture using LSTM, Gated Recurrent Units (GRU), and RNN components for outperforming significantly traditional forecasting models and other non-stacked DNN. Furthermore, they observed that the accuracy of the model does not depend on the size of hidden layers or the number of hidden units. These studies compared the DNN algorithms with other ML models such as XGboost or Random Forest (RF), demonstrating that, on average, DNN models outperform classical approaches. Regarding these studies, among others such as Yousefi et al. (2020), Kadri, Baraoui, and Nouaouri (2019) and Jilani et al. (2019), we can acknowledge the efficacy of this kind of techniques to face the problem of the patient influx to the ED. The studies above mentioned achieve MAPEs between 3.8% and 9.24% for daily aggregations. Nonetheless, it is important to note that comparing different studies can be challenging due to differences in data characteristics for each of them, and the data unavailability.

In addition to DNN that are specifically designed for time series forecasting, Bahdanau, Cho, and Bengio (2014) introduced the Attention Mechanism. This mechanism has revolutionized the field of Deep Learning (DL) in sequence processing, particularly after the development of the Transformers (Vaswani et al., 2017). Attention allows the removal of recurrence and memory cells by processing the complete input data, and learning which points are the most important. After conducting an exhaustive review of ED admissions literature, we have not found any work that uses this mechanism.

Assessing the use of exogenous variables to enhance the forecast, we find several studies in the literature that include factors such as weather and calendar data, most of them presented in Sudarshan et al. (2021) and Jiang et al. (2022). For example, the study conducted by Sudarshan et al. (2021) included weather variables such as temperature, dew point, wind speed and direction, and cloud visibility and coverage. In addition to weather variables, they also included calendar variables such as holidays and the academic calendar. Other studies have used calendar and weather as exogenous variables (Almeida et al., 2022; Fralick, Murray, & Mamdani, 2021; Marcilio et al., 2013; Whitt & Zhang, 2019; Zhang et al., 2022), while some studies only include some of these variables, such as Choudhury and Urena (2020), which only considered calendar variables. Other types of variables that have been used include air quality (Menke et al., 2014; Sun, Heng, Seow, & Seow, 2009), socioeconomic factors (Fralick et al., 2021; Ho et al., 2019; Yousefi et al., 2020) or the flu outbreak level (Xu, Wong, Chin, Wong, & Tsui, 2011). In addition, there have been studies that aimed to evaluate the relationship between ED admissions and information extracted from the Internet, such as Ekström et al. (2015), who employed the number of visits to the healthcare guide in their study area, or Araz, Bentley, and Muelleman (2014), who utilized Google Flu Trends as a predictor.

Our study distinguishes it from existing literature in several aspects. First, we employ the Attention Mechanism for predicting ED patient admissions, which constitutes a contribution compared to the DNN methods used in previous studies. Second, we compare our attention

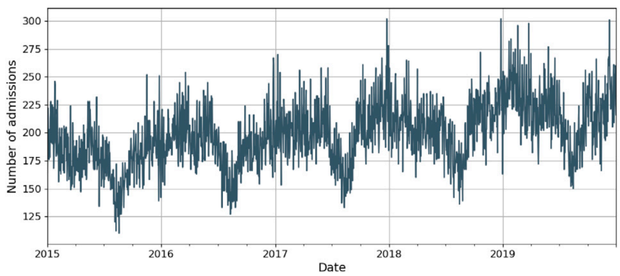

Fig. 1. Patients admissions in HCDGU ED service from January 1, 2015 to December 31, 2019.

model with a more commonly used DNN model in this context, the LSTM-based networks, to determine the better-suited model for our use case. Third, we utilize a wide set of exogenous variables, including weather, calendar, air quality, allergens, and a list of search terms extracted from Google Trends. There is limited research available on some of these variables within this context, such as allergens. Additionally, we conduct an analysis of the independent impact of each type of exogenous variable, which allows us to identify the most important variables among them for ED patient admissions in our hospital. Finally, as this is a context-dependent problem, a new case study adds valuable contributions to the existing literature.

3. Methodology

In this section we present the data used in this study, i.e., ED admissions records and the weather, calendar, allergens, air quality, and Google Trends data. Then, we will present a Naïve model to compare the models used, a neural network model used in previous studies based on LSTM layers, and our proposed DNN based on the Attention Mechanism for accurately forecasting the ED admissions using all the exogenous data.

3.1. Emergency department data

The basis for predicting the patient influx is the historical records of the ED. For our study, we utilized data from the Hospital Central de la Defensa Gómez Ulla (HCDGU) located in Madrid, Spain. Although it is a military hospital, it accepts civilian patients in certain departments, such as the ED. As a result, we analyzed a context where civilians are treated similarly to a public hospital, but with a military designation and organization. Madrid has a population of close to 3.3 million inhabitants and 18 public hospitals disseminated among 605 km^2. Near 80% of the admissions in our study come from the district of Carabanchel (Madrid-South), where the HCDGU is located.

The hospital database used in our study comprises individual records for each patient arriving at the ED. Each record includes attributes such as admission date, gender, age, and Primary Care Center associated. The entire database contains 361 698 patient records from January 1, 2015 to December 31, 2019. To adapt the database for our study, we group the records by the number of daily arrivals to construct our dataset based on the admission series. The raw admission series is depicted in Fig. 1. Conducting an exploratory analysis of the data, we observe significant patterns. There is an increase in the average number of patients treated each year. However, this increase is not uniform, as there are variations over time. For instance, looking at Fig. 2, we can see a clear decrease in admissions during the summer, followed by a rebound until December, which is the month with the highest annual patient average. The distribution of patients by day of the week, presented in Fig. 3, shows a higher influx of patients on Mondays, a midweek drop, and an increase on weekends. The analysis of the distribution of patients by day of the month (depicted in Fig. 4) did

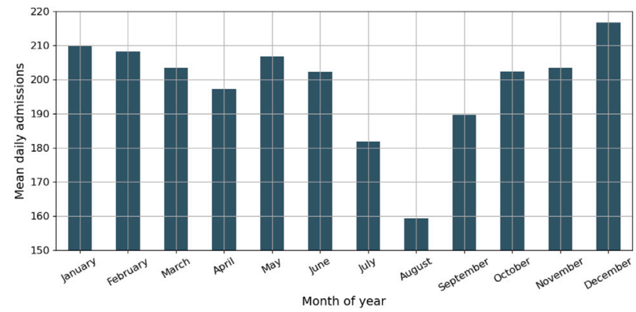

Fig. 2. Distribution of patients by the month of the year.

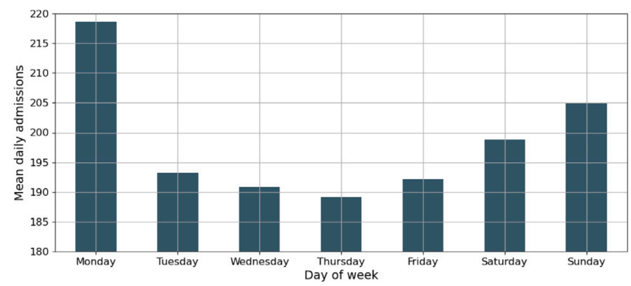

Fig. 3. Distribution of patients by the day of the week.

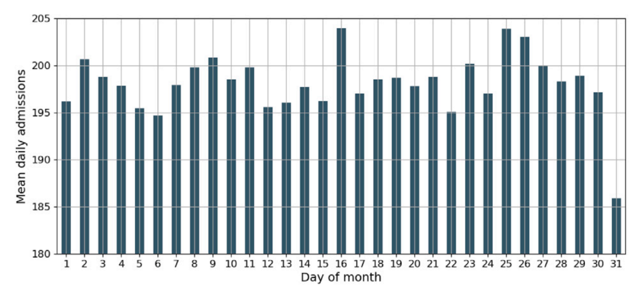

Fig. 4. Distribution of patients by the day of the month.

not reveal significant variability across the different days, except for the 31st day, which may be affected by higher exposure to summer due to its presence in both July and August, as well as the last day of the year, which showed some instability.

Performing an additive decomposition of the time series, we confirmed the previously mentioned trends. The decomposition is presented in Fig. 5. There is a clear weekly seasonality with a maximum variation of around 30 patients (from −9 to 21), an upward trend that appears to have an annual seasonality, and relatively stable residuals throughout the year, but with larger deviations at the end of the year. The annual seasonality is noticeable when comparing the years, as shown in Fig. 6. To provide a more evident comparison, we removed the trend of the series, extracted by the additive decomposition, and normalized the values to range from 0 to 1 for each year, in order to have them on the same scale. The yearly patterns remain relatively consistent, however, there are slight variations in the timing and intensity of the patterns between years. It is important to note the similarities observed between years during the summer months and at the end of the year.

3.2. Exogenous variables

Exogenous variables are those that are not directly related to the target variable (in this case, patient admissions), but may have an indirect

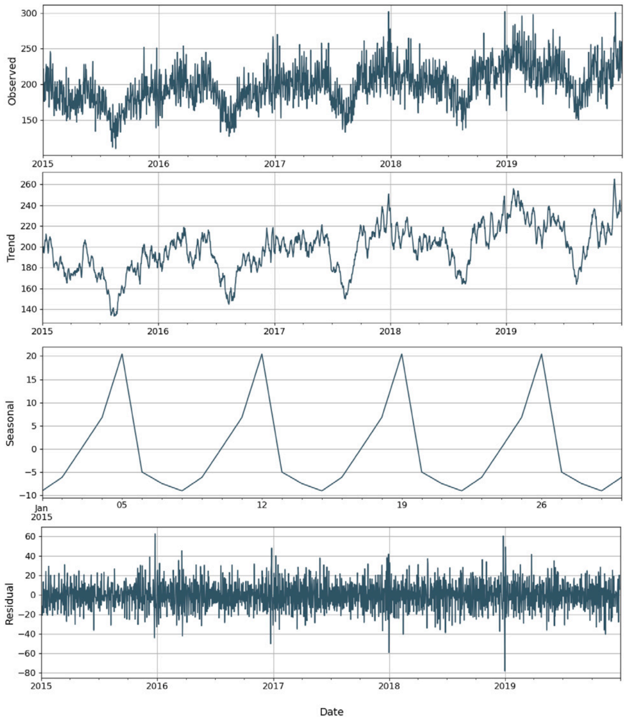

Fig. 5. Decomposition of the admission series from the HCDGU ED service. From top to bottom: raw series, trend, seasonality and residuals.

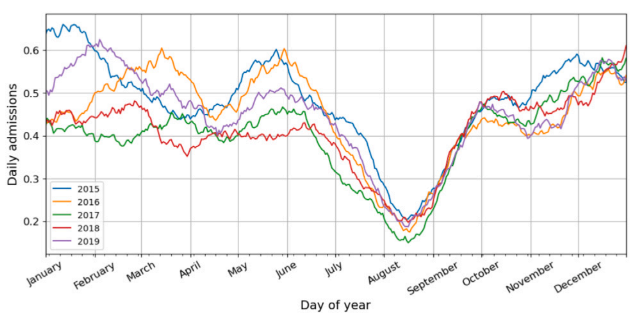

Fig. 6. Additional decomposition of the admission series from HCDGU ED service. The trend of the series was removed and it was normalized by year, applying a rolling window of 30 days to compare the years.

effect on it. They have the potential to provide relevant information for predicting the demand for healthcare services.

In our case, we have decided to introduce the following exogenous variables: weather, calendar, air quality, pollen levels (related to allergens), and Google search trends related to diseases, medicines, or specific events. Each variable is provided as a time-series denoting the relevance of the search term at a specific period. Based on ED staff consulted, these factors can influence the demand for healthcare in various ways; for example, climate can affect the spread of seasonal diseases, while holidays can affect the demand for healthcare due to a higher number of traffic accidents or people traveling away from their usual location. By including these variables, patterns, and relationships that may not otherwise be evident can be identified. Additionally, by taking these variables into account, the model should be able to better adapt to changing situations in the social and physical context in which the hospital is located and, therefore, provide more accurate forecasts.

The calendar variables that have been tested and ultimately selected for our study include weekdays, months of the year, holidays, academic holidays, and weekends. Regarding holidays, we introduced two types

Table 1
Holidays by day of the week for the evaluated period (2015–2019). Academic holidays also contain weekends, but these are not included in the holidays for public work calendar.

Weekday	Labor holidays	Academic holidays
Monday	13	88
Tuesday	10	76
Wednesday	10	78
Thursday	15	77
Friday	15	89
Saturday	6	261
Sunday	4	261
Total	73	1249

of calendars since the labor and academic calendars differ significantly. In Table 1 we can observe that academic holidays contain a higher number of holidays than the labor calendar. For both calendars, we used those specific to Madrid (as in Spain, there are some region-specific holidays). One-hot encoding was used to introduce both the days of the week and the month of the year. The remaining variables are included using a binary approach, with a value of 0 indicating a non-holiday and a value of 1 representing a holiday.

Regarding meteorological data, we have used the daily average temperature (°C), precipitation per day (l/m2), and maximum wind gust speed (m/s). These variables were selected after multiple experiments, as they have shown to provide the best results among all weather variables tested. The data values were obtained through the Spanish State Meteorological Agency (AEMET).[1] Meteorological data can provide valuable information on seasonal trends or weather-related illnesses. Other climate variables that were tested include the direction of the maximum wind gust, average wind speed (m/s), hours of daylight, maximum and minimum daily temperature (°C), daily temperature variation (°C), time of maximum and minimum temperature, and maximum and minimum pressure (hPa).

The air quality variables used in our study are benzene (BEN), methane (CH4), carbon monoxide (CO), ethylbenzene (EBE), non-methane hydrocarbons (NMHC), nitrogen monoxide (NO), nitrogen dioxide (NO2), nitrogen oxides (NOX), ozone (O3), particles less than 10 μm (PM_{10}), particles less than 2.5 μm (fine particles or $PM_{2.5}$), sulfur dioxide (SO2), total hydrocarbons (TCH), and toluene (TOL). These compounds are monitored and regulated by government agencies and environmental control organizations due to their harmful effects on health and the environment. Many of these variables are related to respiratory diseases like asthma or chronic obstructive pulmonary disease (COPD) (Saygın et al., 2017), triggered by different factors such as fuel combustion or other productive activities, e.g., agriculture or construction. These variables are introduced independently into the model, with no aggregation among them. The air quality variables were extracted from Madrid City Council's open data portal.[2] Variables such as metaxylene (MXY), paraxylene (PXY), and orthoxylene (OXY) were also tested, but we decided to exclude them from the study based on prior experimentation.

An additional factor related to respiratory diseases is the presence of allergens. In this respect, we have tested the use of variables such as alnus, alternaria, artemisia, betula, carex, castanea, cupressaceae, fraxinus, gramineae, mercurialis, morus, olea, arecaceae, pinus, plantago, platanus, populus, amaranthaceae, quercus, rumex, ulmus, and urticaceae. After conducting several experiments trying to reduce the dimensionality through Principal Component Analysis (PCA) or manually combining the allergens, we finally decided to introduce all variables to the models. The allergens were extracted from the Spanish

[1] https://www.aemet.es

[2] https://datos.madrid.es/portal/site/egob

Table 2
Summary of all the exogenous variables tested and the selected ones that will be used in our algorithms.

Type of exogenous	All variables tested	Variables selected
Calendar	weekday, month of year, holidays, academic holidays, weekends	weekday, month of year, holidays, academic holidays, weekends
Weather	temperature (avg, min, max, variation, hour max, hour min), max wind direction, wind speed avg, hours of daylight, atmospheric pressure (min, max)	temperature (avg), precipitation, max wind speed
Air Quality	BEN, CH4, CO, EBE, NMHC, NO NO2, NOX, O3, PM10, PM2.5, SO2, TCH, TOL, OXY, MXY, PXY	BEN, CH4, CO, EBE, NMHC, NO NO2, NOX, O3, PM10, PM2.5, SO2, TCH, TOL
Allergens	alnus, alternaria, artemisia, betula, carex, castanea, cupressaceae, fraxinus, gramineae, mercurialis, morus, olea, arecaceae, pinus, plantago, platanus, populus, amaranthaceae, quercus, rumex ulmus, urticaceae	alnus, alternaria, artemisia, betula, carex, castanea, cupressaceae, fraxinus, gramineae, mercurialis, morus, olea, arecaceae, pinus, plantago, platanus, populus, amaranthaceae, quercus, rumex ulmus, urticaceae
Google Trends	allergy, analgesic, antihistamines, appendicitis, asthma, aspirin, bronchiolitis, fall, conjunctivitis, damage, depression, fainting, diabetes, pain, emergency, endometriosis, illness, coping, epilepsy, copd, fever, gastroenteritis, flu, hospital, ibuprofen, incident, heart attack, discomfort, medication, migraine, omeprazole, otitis, paracetamol, prostate, pseudoephedrine, pneumonia, allergic reaction, common cold, brawl, hospital route, bleeding, sinister, high blood pressure, thyroid, urgency, chickenpox, gallbladder	appendicitis, bronchiolitis, conjunctivitis, influenza, hospital, otitis, common cold

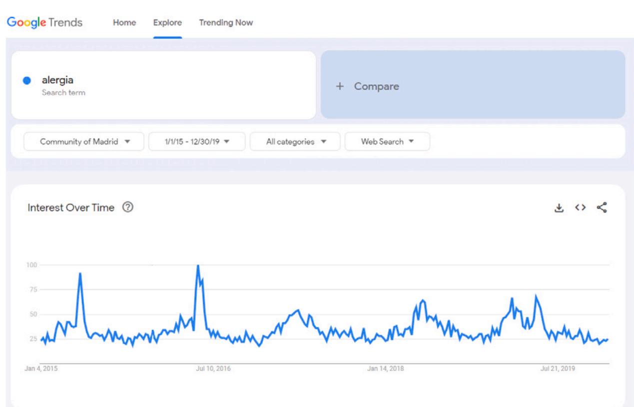

Fig. 7. Example of Google Trends term popularity series. Values for *alergia* (allergy) term for our dataset range dates, and our hospital area.

Society of Allergology and Clinical Immunology (SEAIC),[3] specifically those provided by the station of Clínica Subiza, which is the closest one to the hospital.

To enhance the study of exogenous variables, and in collaboration with ED experts, we decided to utilize the Google Trends tool.[4] This tool extracts the evolution of a specific search term's popularity within a selected range of dates in Google Search. An example of the time series returned for the popularity of a search term can be seen in Fig. 7. The experts aimed to introduce information into the models that would help to monitor and model information, mainly related to diseases. Thus, the initial set of terms was extensive and included the following terms: allergy, analgesic, antihistamines, appendicitis, asthma, aspirin, bronchiolitis, fall, conjunctivitis, damage, depression, fainting, diabetes, pain, emergency, endometriosis, illness, coping, epilepsy, COPD, fever, gastroenteritis, flu, hospital, ibuprofen, incident, heart attack, discomfort, medication, migraine, omeprazole, otitis, paracetamol, prostate, pseudoephedrine, pneumonia, allergic reaction, common cold, brawl, hospital route, bleeding, sinister, high blood pressure, thyroid, urgency, chickenpox, and gallbladder. However, our experiments have revealed that utilizing a large set of terms in this manner requires significant time for training the models. As a result, we relied on the expertise of our ED professionals to narrow down the list to those illnesses that are frequently observed in patients arriving at the ED service. Therefore, the final list was reduced to appendicitis, bronchiolitis, conjunctivitis, influenza, hospital, otitis and common cold. It should be noted that the search terms in Google Trends were introduced in Spanish due to the context of our hospital. Table 2 summarizes all exogenous variables tested and the ones selected for this study.

3.3. Naïve baseline forecasting model

For having a baseline, we will employ a naïve model based on value persistence to establish a threshold that indicates whether the forecasting models perform adequately. For the Naïve model, the last values are replicated to make the forecast in the time horizons studied. For a 1-day horizon, the value from the previous day is replicated, while for a 3-day horizon, the values from the last 3 days are replicated, and so on for a weekly horizon. Eq. (1) describes the implementation of the Naïve model, being $\hat{y}_t$ the predicted admission value at day t and y_{t-i} the admission record at day $t-i$. For aggregated time periods of 3 and 7 days, the value from the previous aggregated period is replicated to make the prediction, as in Eq. (2).

$$\hat{y}_t = y_{t-i} \tag{1}$$

$$\hat{y}_{t_{agg}} = \sum_{i=1}^{j} y_{t-i} \quad \text{for } j \text{ days} \tag{2}$$

3.4. LSTM baseline forecasting model

In addition to the Naïve model described previously, we have implemented a state-of-the-art method that uses RNN for the same

[3] https://www.polenes.com
[4] https://trends.google.com

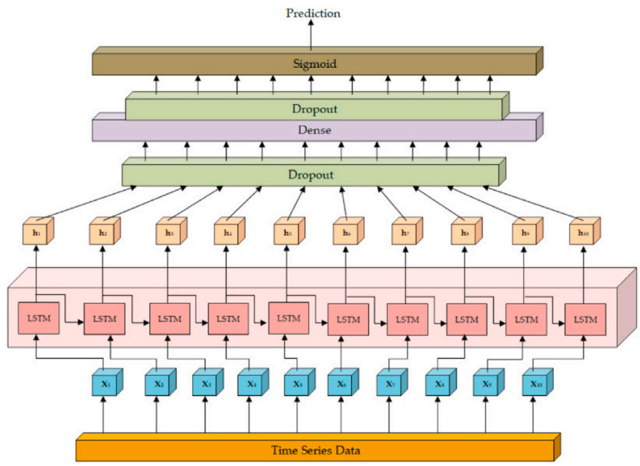

Fig. 8. LSTM-based model proposed by Sudarshan et al. (2021).

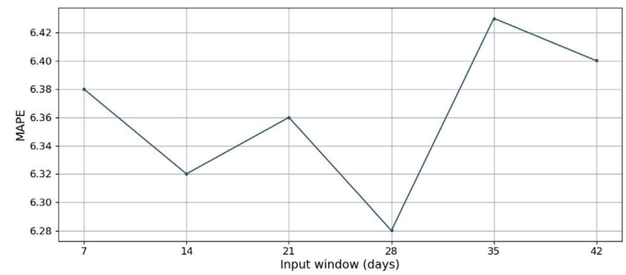

Fig. 9. MAPE evolution depending on the size of the input length in days.

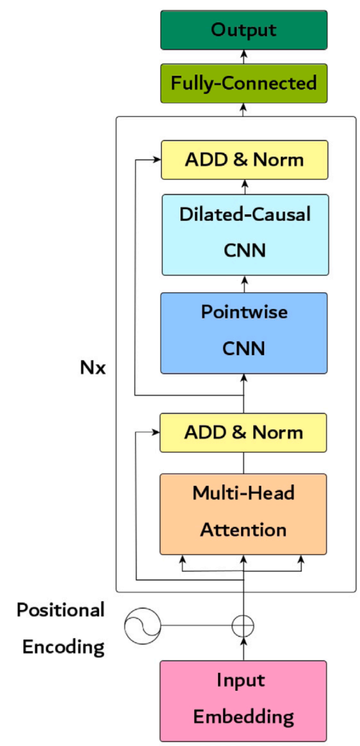

Fig. 10. DNN proposed model using attention layers.

domain application. Particularly, we will use an LSTM model as proposed by Sudarshan et al. (2021) and depicted in Fig. 8. This model outperformed other models used in their study, such as RF and CNN, making it a good candidate to compare the performance with our Attention-based architecture. The model implements an LSTM with 75 length hidden size, whose output passes through two dropout layers to avoid overfitting issues. Then, a dense layer using a sigmoid activation function flattens the output to the desired forecasting horizon. The hyperparameters used in this model are directly extracted from the work of Sudarshan et al. (2021). The model uses the previous 10 days as an input and provides 3 or 7-day forecasts. As well, the model is fed with some exogenous variables including weather and calendar information. Specifically, they include: wind speed, visibility range, temperature, dew point, cloud coverage, day of the week, and holidays.

We have implemented the same LSTM-based model with two variations to enable a fair comparison with our Attention-based model: we have trained the assessed models with the same sets of exogenous variables, and we have used the same input length of the historical records. For the first modification, we have trained various models: one without exogenous data (only using the admission records), and others in which we have included the calendar data, weather data, allergens concentration, air quality measures, and Google Trends information in the same way as we have done with our proposed architecture.

To assess what is the best input length to feed the network, we have conducted an experiment to determine such value. We know that our ED records present a strong weekly seasonality, thus we have trained this model with different input lengths that are multiple of a week, including the admissions records and the calendar information. The results of this experiment are depicted in Fig. 9. From these results, we can see that the best MAPE is obtained when the input sequence has a length of 28 days (4 weeks). Therefore, all the models presented will use such value as input length. This is consistent with our exploratory analysis of the data, in which we also see a clear monthly seasonality. As well, we can see that the worst values are achieved when we use more than a natural month as the input.

3.5. *Attention forecasting model*

Recently, the design of DNN architectures has been greatly influenced by the use of the Attention mechanism proposed by Bahdanau et al. (2014). This allows the model to consider which parts of the input information are more important, using the weight assigned to the relationships between the input data while reducing the network dimensionality. The transformers proposed by Vaswani et al. (2017) contributed to the increasing popularity of this mechanism.

For our ED admissions forecasting, we rely on a transformer model using multi-head attention for processing the multivariate time series input, i.e., the admissions records plus the exogenous variables. We will have the same input as in the LSTM model: 28 days of admission records plus the vector with the exogenous variables. The DNN proposed is formed by a positional encoding connected to the input embedding, followed by the attention mechanism, a Pointwise CNN, a Dilated-Causal CNN, and a normalization step previous to the output flatten of the forecast, as depicted in Fig. 10. In the following, we will briefly discuss each step of our DNN.

In the original work proposed by Vaswani et al. (2017) it is necessary to include temporal or positional information in the input sequence before feeding it into the network. The aim is to generate a distinct representation for every input by providing the original pre-trained embedding with positional information based on sine and cosine functions. Due to the nature of the time series data, some research has suggested a learnable positional encoding as a more flexible approach for capturing temporal relationships within the data (Kazemi et al., 2019; Zerveas, Jayaraman, Patel, Bhamidipaty, & Eickhoff, 2021). This method requires to introduce an embedding layer in the transformer structure that learns embedding vectors for each position index in conjunction with other model parameters.

In accordance with our experimental findings and results, we have chosen to adopt an approach that is similar to the original fixed positional encoder with the addition of seasonality information as presented in previous sections. We achieve this by concatenating positional information related to the day of the week and month of the year at each time step. Consequently, we define two dimensionalities, d_t for the embedding and d_p for the position, with a total dimensionality of $d = d_t + d_p$.

The processed input is fed to the transformer where we apply a scaled dot-product attention to determine the existing correlation between our model inputs irrespective of the distance between them. The original input is projected into a query vector, denoted as Q, with a dimensionality of d_q, a key vector, K, with a dimensionality of d_k and a value vector, V, with a dimensionality of d_v, the scalar dot-product attention used is given by Eq. (3).

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

In our particular case, we will work with a multi-head attention approach: the vectors Q, K, and V are projected H times, being H the number of heads. Each head performs the attention function in parallel, concatenating the obtained values and projecting them one last time, allowing the model to attend to information in different subspaces at the same time. This is presented in Eq. (4). It is in those subspaces where we expect that the exogenous variables have more impact in the training phase of the model. The information extracted from the attention layers is normalized and added to the raw input data prior to being processed by the next layer.

$$\begin{aligned} MultiHead(Q, K, V) &= Concat(head_1, \dots, head_H)W^O \\ \text{where } head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{4}$$

Existing transformer models weigh the outputs of the attention layer using a fully connected neural network. In our case, we take advantage of a pointwise convolutional layer instead, which offers an advantage when interpreting two-dimensional multivariate time series inputs, including time domain features and variables. It is a variation of the CNN with filter size 1, similar to a fully connected network, multiplying a value by weight, but keeping the input shape and adjusting the feature dimensions through the number of convolution filters.

The output of the pointwise CNN is processed by a dilated-causal CNN that can be useful for extracting time series characteristics. The dilated convolution method is a variation of the 1D-convolutional operation that applies the filter by neglecting certain inputs, and finding patterns in wider time frames. The formula that defines the dilated convolution operation is depicted in Eq. (5), where F indicates the input feature, k is the convolution filter, l is the dilation rate (that defines how much we want to widen the kernel) and p, s, and t refers the position of the features. Causal convolution assumes that current data is only affected by past data, instead of taking into account all adjacent data in both directions (normal convolution), i.e., we focus on the past information. Combining these methods allows us to extract compressed features from a given long time series input.

$$(F *_l k)(p) = \sum_{s+l_t=p} F(s)k(t) \tag{5}$$

Finally, the information provided by the dilated-causal CNN is normalized and added with the attention layers and passed through a dense layer to flatten the vector to the desired output forecast, that is, 1, 3, or 7 days.

To achieve the best performance, we have tested several hyperparameters:

- For training and validating the model we have assessed various metrics: MSE, MAE, RMSE, and Hubber Loss. In our experiments, we used MSE as it has the best behavior.
- For the batch size we evaluated {32, 64, 128} deciding to use 32.
- The following optimizers were tested: Adam, RMSprop, Adadelta, AdamW, Adagrad and AdaBelief. We finally chose Adam using learning rates in the range $[10^{-5}, 10^{-1}]$, β_1 in [0.7, 0.9], β_2 in [0.7, 1.0] and ϵ in $[10^{-5}, 10^{-9}]$. The selected values are 0.9, 0.98 and 10^{-9} for β_1, β_2 and ϵ respectively.
- For the positional encoding we have evaluated {32, 64, 128, 256} units and various dropout rates. The best results are obtained with two layers with 256 units and 0.3 dropouts.
- For the multi-head attention we have tested the number of heads in the set {2, 4, 6, 8, 10} and head sizes of {16, 32, 64, 128, 256, 512}, being the ones that achieve the best results 4 heads with a size of 128.
- For the convolutional layers of the feedforward block we assessed different numbers of kernels {32, 64, 128, 256, 512}. The best results were achieved with 64 kernels.
- To avoid overfitting we have added dropout in our convolutional and attention layers, trying values of {0, 0.1, 0.2, 0.3, 0.4}. Based on our experiments, there are no overfitting when no dropout are used. Therefore, we set the dropout to 0.
- Finally, we have to set the number of iterations for the transformer. We have evaluated {2, 4, 6, 8, 10, 12} and selected 4 iterations. The higher number of iterations notably increase the execution time while there are no significant improvement in the results.

It is worth mentioning that, while the proposed attention model has more layers than the LSTM presented in the previous section, the attention model has 127 561 parameters meanwhile the LSTM has 2 125 101 parameters for the models that do not consider exogenous variables. The models that use calendar information have 445 481 and 2 131 701 parameters for the attention model and the LSTM models respectively. This has implications in both the training and operation phase. While the training requires more time in the case of the attention model, it provides better results and its execution requires less time than the LSTM model.

3.6. Experimental settings

As is commonly done in ML projects, we divide our dataset into three sets: train, validation, and test. The training set contains the 68% of the historical records, which corresponds to 1246 records from January 1, 2015, to May 30, 2018. The validation set comprises 12% of the data, that is, 215 records from May 31, 2018, to December 31, 2018. Finally, the test set includes the entire year 2019, that is, the 20% of the data (365 records) from January 1, 2019, to December 31, 2019. Table 3 presents a summary of the split.

Table 3
Split of the historical records used in our work.

Partition	Percentage	Days	Period of time (From-To)
Train	68%	1246	01/01/2015–30/05/2018
Validation	12%	215	31/05/2018–31/12/2018
Test	20%	365	01/01/2019–31/12/2019
Total	100%	1826	01/01/2015–31/12/2019

We attempt to use cross validation by creating large partitions to ensure that we can capture both the trend and the weekly pattern of our data. This results in a limited number of partitions and makes the models unable to capture the yearly seasonality. Similarly, we tried to use expanding or sliding windows, but we faced similar problems. Therefore, we retain the entire dataset.

To enhance the learning procedure of the selected algorithms within this research, the time series values are subjected to a min–max normalization, ranging from 0 to 1, prior to their incorporation into the models. As we commented in Section 3.5, the batch size of the input data was set to 32, the loss function was MSE, and the Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of 0.001, reducing this value when MSE validation loss stopped improving with a reduction factor of 0.6 and a patience number of epochs of 50. Furthermore, in order to avoid overfitting, we use the early stopping technique to finish the process when there is no improvement in the validation set for more than 100 epochs.

3.7. Evaluation metrics

For evaluating the proposed models, we use the MAPE, R^2, Mean Bias Error (MBE), and Pearson Correlation (ρ). The definitions of these metrics are presented below:

- MAPE is the mean of all absolute percentage errors between the predicted and actual values, returning the error as a percentage and computed as in Eq. (6). It is useful to compare model accuracy across use cases and datasets.
- R^2 or the coefficient of determination is the percentage of variation explained by the relationship between two variables, in our case, predicted and actual values. It provides a measure of how well-observed outcomes are replicated by the model. It is computed as in Eq. (7).
- MBE is the mean of the difference between the predicted values and the actual values. A positive bias means the error from the data is overestimated and a negative bias means the error is underestimated. We use it as a percentage, computed as in Eq. (8).
- ρ measures the linear relationship between the predicted values and the actual values. Like other correlation coefficients, it varies between −1 and +1, being +1 when the predicted and actual values are the same. It is computed as in Eq. (9).

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{7}$$

$$\text{MBE (\%)} = \frac{100}{n \cdot \bar{y}}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{8}$$

$$\rho = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{9}$$

4. Results and discussion

In this section, we present the results obtained by the different evaluated models. First, we show the metrics assessment and then we discuss how our proposal can be used to enhance the effectiveness of the ED service.

4.1. Results

In our experiments, we have conducted a forecast with the three presented models, i.e., Naïve, LSTM, and Attention. The three models have been executed to obtain the admissions predictions for the test period (the whole 2019 year) on three forecasting horizons: 1, 3 and 7 days. The 1-day and 3-day horizons are used for a methodological comparison with Sudarshan et al. (2021), while the 7-day horizon is used to evaluate the performance of the models over a longer forecasting period, which is relevant for the weekly planning conducted by the managers of the service. The Naïve model only considers the historical admissions record while the DNN models also allow for exogenous variables. In this case, the LSTM and Attention models have been tested with calendar, weather, air quality, allergens, and Google Trends information separately to assess the impact of those variables in the forecast.

The results for the 1-day horizon are presented in Table 4. It shows that the DNN-based models (LSTM and Attention) outperform the Naïve model in all metrics. Between the DNN models studied, the Attention models achieve better performance compared to LSTM models for all exogenous variables, also being less biased as assessed by the MBE. Specifically, the Attention model with calendar data exhibits better performance, with a MAPE of 5.98%, R^2 of 0.64, and ρ of 0.80, surpassing the LSTM models with the same exogenous variable with a MAPE of 6.32%, R^2 of 0.58 and ρ of 0.77. These differences are consistent across the other exogenous variables for each of the networks. It is worth noting that the models with the calendar data improve upon the base models without exogenous data, while for the allergen, climate, and air quality variables they remain very close to the base models. In contrast, the models trained with Google Trends variables achieve lower metrics than the base models and show a higher bias than the other models. It should be noted that all models present a high correlation (ρ) with a value range from 0.72 to 0.80.

Table 4
Metrics for 1-day predictions, ordered from best MAPE to worst. All Pearson correlations are significant ($p_{value} < 0.05$).

Network	Exogenous	MAPE (%)	R^2	MBE (%)	ρ
Attention	Calendar	5.98	0.64	−0.45	0.80
Attention	Air quality	6.23	0.58	−1.38	0.78
Attention	Weather	6.27	0.59	−1.49	0.78
Attention	Allergens	6.30	0.58	−1.81	0.78
Attention	–	6.31	0.59	−1.09	0.77
LSTM	Calendar	6.32	0.58	−1.92	0.77
LSTM	Weather	6.55	0.54	−2.37	0.76
LSTM	Air quality	6.63	0.52	−0.52	0.74
Attention	Google Trends	6.66	0.52	−3.11	0.76
LSTM	Allergens	6.69	0.53	−2.85	0.76
LSTM	–	6.96	0.51	0.95	0.72
LSTM	Google Trends	7.07	0.46	−3.63	0.73
Naïve	–	9.02	0.23	0.16	0.62

Table 5
Metrics for 3-days predictions, ordered from best MAPE to worst. All Pearson correlations are significant ($p_{value} < 0.05$).

Network	Exogenous	MAPE (%)	R^2	MBE (%)	ρ
Attention	Calendar	6.05	0.63	−0.58	0.80
Attention	Air quality	6.21	0.59	−1.58	0.78
Attention	Weather	6.27	0.59	−1.77	0.78
LSTM	Calendar	6.28	0.59	−2.11	0.78
Attention	Allergens	6.30	0.59	−2.07	0.78
Attention	–	6.35	0.58	−1.34	0.77
LSTM	Weather	6.53	0.54	−2.54	0.76
LSTM	Air quality	6.57	0.53	−0.65	0.75
LSTM	Allergens	6.67	0.53	−3.04	0.76
Attention	Google Trends	6.70	0.51	−3.39	0.76
LSTM	Google Trends	7.03	0.46	−3.84	0.74
LSTM	–	7.22	0.48	1.10	0.70
Naïve	–	11.01	−0.03	0.08	0.49

The results for the 3-day forecasting, presented in Table 5, show that the Attention model with the calendar data still maintains the best results, with a MAPE of 6.05%, R^2 of 0.63 and a ρ of 0.80. The rest of the models seem to remain similar, except for the LSTM base model, which falls to a MAPE of 7.22%. For this case, the LSTM model with calendar data slightly outperforms the base Attention model with a performance of 6.28% of MAPE, 0.59 of R^2, and ρ of 0.78. Both architectures significantly outperform the Naïve model, which stands at 11.01% of MAPE, R^2 of −0.03, and ρ of 0.49. In this case, the correlation presented by the models is in the range from 0.80 for the best model to 0.70 for the base LSTM model. The LSTM models also present a larger bias for each exogenous data in comparison to the Attention models.

When assessing the forecast at 7 days, we can notice that the distance between DNN models slightly increases. This can be seen in Table 6. While the base LSTM model has a MAPE of 7.41% and R^2 of 0.45, the base Attention model obtains a MAPE of 6.38% and R^2 of 0.57. When comparing the predictions made at 1 and 3 days, it can be observed that the difference between the two models grows as the

Table 6
Metrics for 7-days predictions, ordered from best MAPE to worst. All Pearson correlations are significant (p_{value} < 0.05).

Network	Exogenous	MAPE (%)	R^2	MBE (%)	ρ
Attention	Calendar	6.10	0.63	−0.59	0.79
Attention	Air quality	6.23	0.58	−1.61	0.78
Attention	Weather	6.32	0.58	−1.89	0.77
LSTM	Calendar	6.33	0.58	−2.21	0.78
Attention	Allergens	6.36	0.58	−2.14	0.78
Attention	–	6.38	0.57	−1.27	0.76
LSTM	Weather	6.57	0.53	−2.62	0.76
LSTM	Air quality	6.58	0.53	−0.62	0.74
Attention	Google Trends	6.72	0.51	−3.45	0.76
LSTM	Allergens	6.75	0.53	−3.10	0.76
LSTM	Google Trends	7.11	0.45	−4.02	0.73
LSTM	–	7.41	0.45	1.27	0.68
Naïve	–	9.58	0.19	−0.29	0.59

prediction horizon is extended. Incorporating exogenous variables in the models has the potential to decrease the performance gap between the two models. However, Attention-based models still outperform LSTM-based models in each type of exogenous variables, obtaining MAPEs of 6.10% for the calendar, 6.23% for air quality, 6.33% for allergens, and 6.72% for Google Trends, while LSTM models remain at 6.33%, 6.58%, 6.57%, and 7.11%, respectively. Once more, we confirm that in this instance both models exhibit significant improvement in performance over the Naïve model, which slightly improves regarding to the 3-days forecast due to the weekly seasonality of the series. Regarding the correlation, the range increases again from 0.79 for the best model to 0.68 for the base LSTM model. Also, the LSTM models exhibit, again, a larger bias for each exogenous data compared to the Attention models.

In addition to the daily prediction results, we have also made predictions for the aggregated periods of 3 and 7 days. Table 7 shows the metrics obtained for the 3-day aggregated prediction. In this case, the Attention models outperform the LSTM models for every independent variable. Furthermore, the model that uses calendar variables obtains the best results (4.00% of MAPE and 0.75 of R^2). Also, the Attention model with weather variables outperforms the Attention model with air quality variables, achieving values of 4.32% of MAPE and 0.71 of R^2, compared to 4.36% and 0.68, respectively. It should be noted that the Attention model with allergens is not able to improve the performance of the base Attention model. The 7-day aggregated prediction confirms that with Attention, weather information improves the air quality model, as shown in Table 8. For this case, the 7-day aggregated predictions, the metrics improve to 3.19% of MAPE, 0.80 of R^2, and a ρ of 0.90 for the best model (being the Attention model with calendar variables the best again). As for the best LSTM model, it achieves a MAPE of 3.61%, R^2 of 0.72, and a ρ of 0.87. In the aggregated forecasts, both models outperform the Naïve model. However, for the 7-day horizon, the Naïve model, taking advantage of weekly seasonality, comes close to the results obtained with the LSTM model using Google Trends, even surpassing it in R^2. For the aggregated forecasts the correlation is stronger than single-day models, achieving 0.90 for the best model.

To determine which type of exogenous information aids in improving the predictive accuracy of the models, we have computed a score based on the sum of MAPEs obtained for different time horizons and DNN models. This approach facilitates the overall comparison among the different exogenous variables, including base models without any exogenous variable. As the MAPE metric is based on the error percentage, lower values mean better model performance. Thus, for our exogenous score, lower values also mean better overall performance, but in this case, it should be noted that it is not related to the error percentage of any of them, being a qualitative more than quantitative comparison among models. The results of this score for each exogenous variable are depicted in Fig. 11. We observe that the calendar variables

Table 7
Metrics for 3-days aggregated predictions, ordered from best MAPE to worst. All Pearson correlations are significant (p_{value} < 0.05).

Network	Exogenous	MAPE (%)	R^2	MBE (%)	ρ
Attention	Calendar	4.00	0.75	−0.52	0.87
Attention	Weather	4.32	0.71	−1.70	0.86
Attention	Air quality	4.36	0.68	−1.52	0.84
Attention	–	4.37	0.70	−1.26	0.85
LSTM	Calendar	4.38	0.69	−2.05	0.85
Attention	Allergens	4.61	0.68	−1.99	0.84
LSTM	Weather	4.67	0.63	−2.49	0.82
LSTM	Air quality	4.84	0.62	−0.59	0.80
Attention	Google Trends	4.89	0.61	−3.34	0.84
LSTM	Allergens	5.08	0.61	−2.97	0.83
LSTM	–	5.20	0.61	1.10	0.79
LSTM	Google Trends	5.42	0.50	−3.79	0.79
Naïve	–	8.59	0.25	0.19	0.63

Table 8
Metrics for 7-days aggregated predictions, ordered from best MAPE to worst. All Pearson correlations are significant (p_{value} < 0.05).

Network	Exogenous	MAPE (%)	R^2	MBE (%)	ρ
Attention	Calendar	3.19	0.80	−0.59	0.90
Attention	Weather	3.29	0.76	−1.89	0.89
Attention	Air quality	3.36	0.73	−1.61	0.87
Attention	–	3.54	0.74	−1.40	0.87
LSTM	Calendar	3.61	0.72	−2.21	0.87
LSTM	Weather	3.84	0.65	−2.62	0.85
Attention	Allergens	3.91	0.71	−2.14	0.87
LSTM	Air quality	4.06	0.66	−0.62	0.83
Attention	Google Trends	4.23	0.64	−3.45	0.87
LSTM	–	4.43	0.64	1.27	0.81
LSTM	Allergens	4.60	0.63	−3.10	0.86
LSTM	Google Trends	4.90	0.48	−4.02	0.80
Naïve	–	5.32	0.53	−0.35	0.76

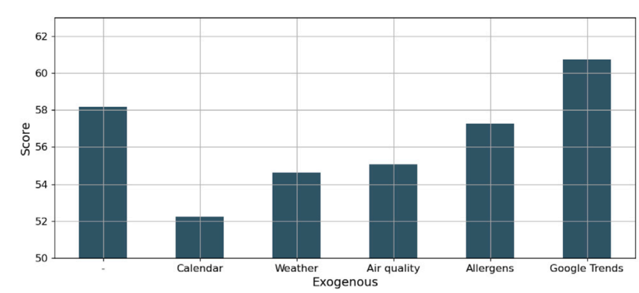

Fig. 11. Exogenous score (lower values are better). All variables generally improve the DNN base models, except Google Trends data, being the calendar the most useful data for the models assessed.

seem to provide the most useful information, with a score of 52.24, improving upon the model without exogenous variables, which has a score of 58.17. We observe that the calendar variables seem to provide the most useful information, with a score of 52.24, improving upon the model without exogenous variables, which has a score of 58.17. The weather and air quality variables both have very similar scores, with values of 54.63 and 55.07 respectively. The allergens variable also scores lower than the model without exogenous variables, with a score of 57.27. Lastly, the Google Trends variables seem to be penalizing the models, with a score of 60.73.

4.2. Discussion

According to the findings presented in the previous section, we extract some relevant information to be discussed. It is important to highlight that the inclusion of exogenous variables shows a significant improvement in both DNN models compared to the baseline, as they were able to identify relevant patterns in the training data and provide

Table 9
Summary table comparing the best results among our Attention-based model with the LSTM-based model adapted from (Sudarshan et al., 2021) and the Naïve model for all studied forecasting horizons.

Model	Prediction horizon									
	1-day		3-day		3-day agg		7-day		7-day agg	
	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2
Attention	5.98	0.64	6.05	0.63	4.00	0.75	6.10	0.63	3.19	0.80
LSTM	6.32	0.58	6.28	0.59	4.38	0.69	6.33	0.58	3.61	0.72
Naïve	9.02	0.23	11.01	−0.03	8.59	0.25	9.58	0.19	5.32	0.53

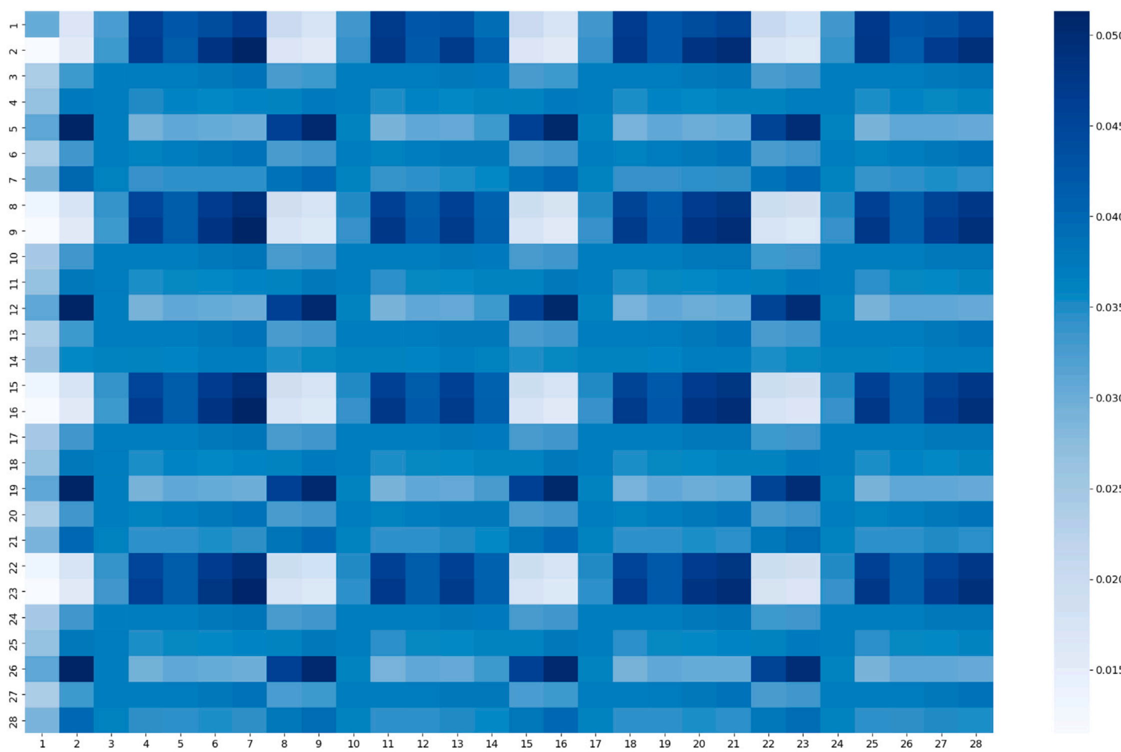

Fig. 12. Self attention patterns for the 28-day input window.

more accurate forecasts for future scenarios. This validates the functionality of the models for their application in predicting the patient's influx to the ED. Taking into account the high values of ρ for all cases, where values above 0.75 were obtained in almost all cases and up to 0.90 in aggregated series, we can infer that the models are producing accurate and reliable predictions. Consequently, it seems that our models are providing valuable results in understanding social behavior in the area in relation to the ED.

According to our experiments, Attention models outperform LSTM models for all cases as shown in Table 9. Further analyzing the Attention model, we have observed that it is able to accurately capture the weekly pattern. In Fig. 12, the scores of one of the attention heads to each input in the window are represented by different colors according to the assigned importance (learned by the network). It can be observed that for each of the 28 days in the input window, a different score is assigned to the rest of the days. The weekly pattern is evident as the assigned score is similar every seven days. Thus, four cycles of the pattern are formed coinciding with the number of weeks in the input, both horizontally and vertically in the graph.

Analyzing the models' performance and the differences between them, some differences can be observed in the patterns extracted when visualizing the predictions made by each model. Fig. 13 shows the 1-day forecasts for the best Attention-based model (using calendar data) and Fig. 14 shows the LSTM-based model forecasts for the same case. Those figures depict the forecast for the whole test set (year 2019). By its side, Figs. 15 and 16 depict the 7-days forecasts without aggregations for the best Attention and LSTM models. The LSTM-based models show a tendency to have a relatively stable prediction around the mean, adjusting the weekly pattern according to the yearly trend. In contrast, the Attention-based models better capture the variability of the observation series more effectively than the LSTM-based models. Fig. 17 shows more clearly the smoothing of the LSTM-based model compared to the greater variability in the adjustment of the Attention-based model. We should emphasize that the months of January and February in our testing set appear to slightly break the annual seasonality of the series, causing the metrics of both models to decrease. Notwithstanding, even in the worst case, the Attention model is able to achieve accurate predictions for some admissions peaks (being a larger error of less than 20 patients) where the LSTM clearly underestimates the admissions, with errors larger than 30 patients.

Regarding the exogenous variables we have observed that the calendar enhances the baseline model in both scenarios, indicating the significance of this variable in comprehending the patterns of ED admissions, as it was reflected with the exogenous score. Weather and air pollution have a positive impact on the model but not with the same importance as calendar data, improving the base model in all cases and reflecting it in the exogenous score. For allergens, although the base model sometimes improves by adding them, they often introduce noise, making it difficult to recognize patterns in the series and having a limited impact on the model. Finally, the inclusion of Google Trends information into the model seems to worsen the performance of the baseline model by introducing too much noise that

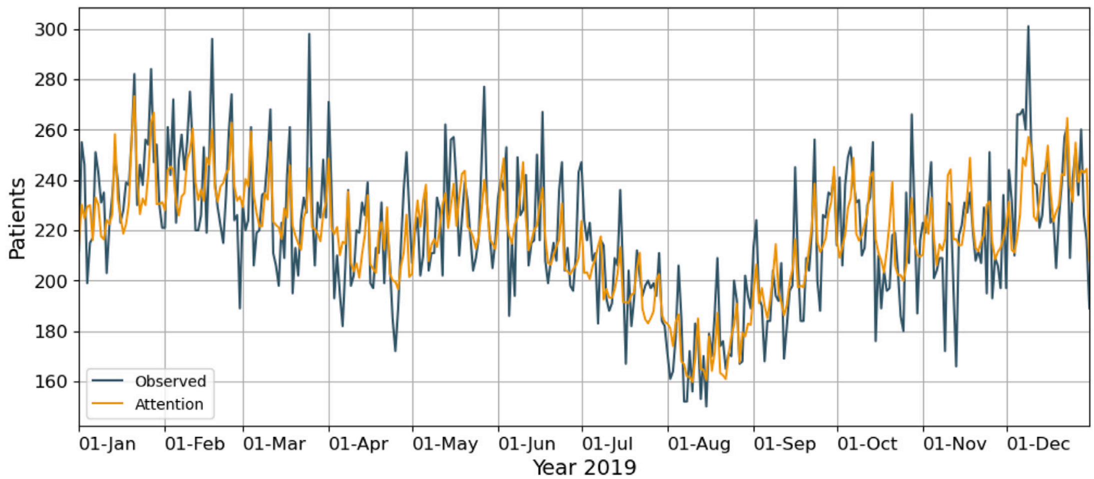

Fig. 13. Attention model 1-day forecasts from the best model (using calendar data).

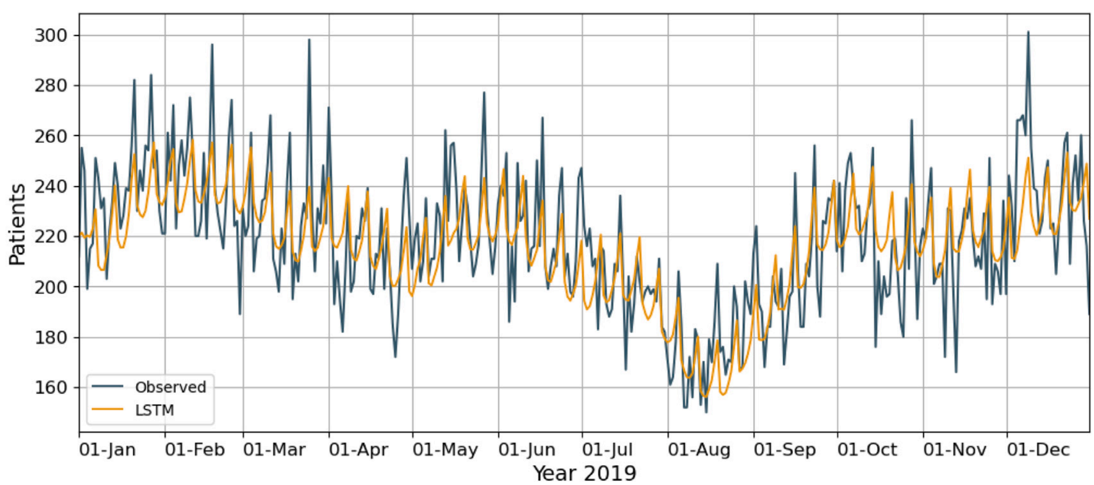

Fig. 14. LSTM model 1-day forecasts from the best model (using calendar data).

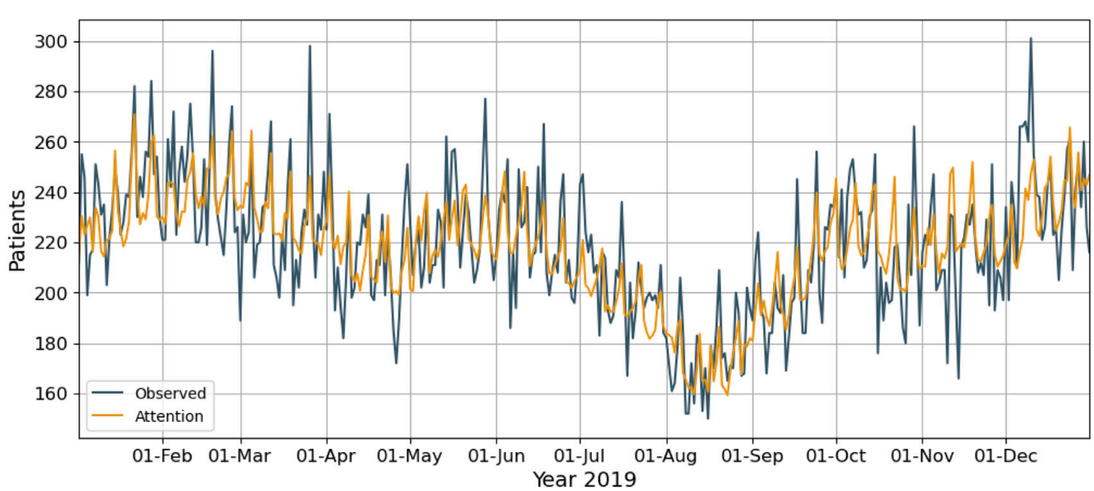

Fig. 15. Attention model 7-day forecasts from the best model (using calendar data).

causes useful information from the series to be lost. This can be seen in Fig. 18 where the attention scores learned by the network for the Google Trends information. Therefore, proper selection of exogenous variables is crucial to achieve good performance in forecasting models.

Finally, we tried to include all exogenous variables in our proposed Attention network. We do not offer the results here because they were worse than those presented, and in some cases even worse than the Naïve model. Our assessment reveals that the network is not able

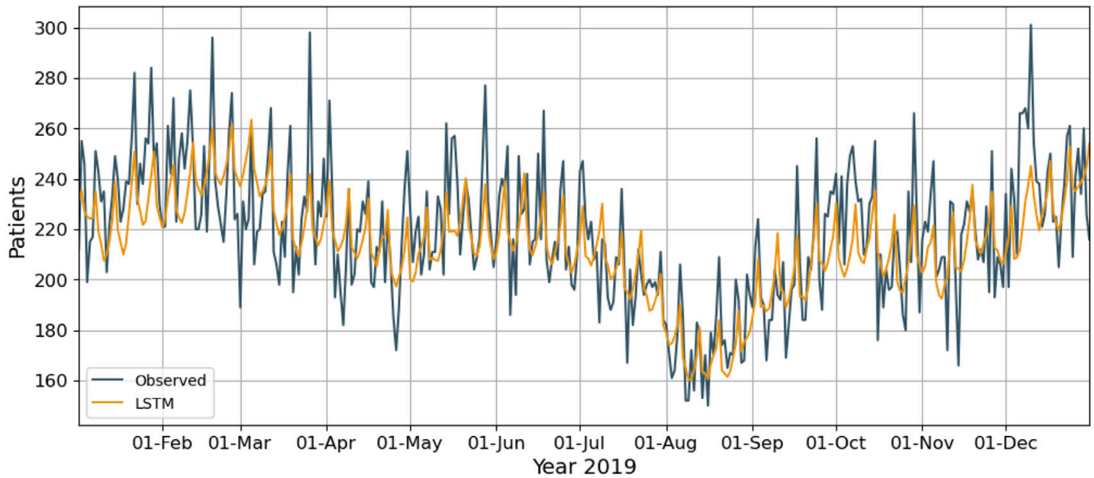

Fig. 16. LSTM model 7-day forecasts from the best model (using calendar data).

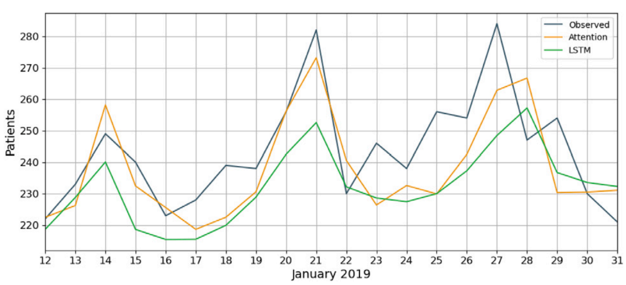

Fig. 17. 1-day forecasting predictions comparison of LSTM and Attention models for January 12 to 31, 2019.

to focus adequately on the relevant features when dealing with such a large amount of information. Moreover, there are some exogenous variables that seem to provide similar information (e.g. air quality and weather), which makes the network focus its attention on these variables and neglect more relevant ones (e.g. calendar). To obtain the best accuracy it is required to perform a combination of all variables of the hospital under study. However, such an approach requires a lot of time to train all possible models. In order to keep the article concise, we have tried to provide an assessment of the individual variables so further studies can focus on variable combinations.

In the field of resources ED management, accurate prediction of the demand is essential to ensure that there are enough healthcare professionals available to attend to the patients. To achieve more accurate predictions, models have been developed that use aggregated data to predict the demand for three days and seven days. These models have been shown to significantly improve forecast accuracy. Although these models are not useful for determining patient demand for a specific day, they can be very useful for sporadic recruitment and for planning specific time periods, such as holidays that are expected to have a high demand for emergency services. Furthermore, overestimating demand can have serious economic consequences. If resources are planned for a higher demand than what is actually required, it may be necessary to pay staff more to cover the anticipated demand, resulting in significantly higher costs than what was budgeted for. This can lead to a difficult economic situation for the hospital, especially if it occurs frequently. To avoid this problem, it is important to have accurate and reliable models to predict healthcare demand. The Attention-based models presented here for our context have been shown to have lower bias compared to those obtained using LSTMs, suggesting that they are a valuable tool for improving human resource planning and avoiding costs associated with overestimation.

In the field of ED forecasting admission, the results of using a model are highly dependent on the context in which they are applied. For this reason, we believe that it is important to adapt the models used in the literature to the specific context of the problem being addressed to achieve an accurate comparison between predictions and their metrics. In this regard, we adapted the LSTM-based model from Sudarshan et al. (2021) and applied it to our dataset to make a precise comparison between our model's forecasts and the corresponding metrics. This adaptation allowed us to obtain more accurate and realistic results in our hospital context.

5. Conclusions

In this study, an Attention-based DNN model has been developed to forecast patient admissions in the ED by using different types of exogenous variables to improve the model's accuracy. The results of the Attention-based model have been compared with a DNN model adapted from Sudarshan et al. (2021) and a Naïve baseline to evaluate the effectiveness of our approach. The experimentation indicates that both models are adequate, with the Attention-based model outperforming the LSTM-based model for all cases. Among the considered exogenous variables, we found that the calendar provides the most useful input, followed by weather and air quality. In contrast, Google Trends data confuses the models and reduces their effectiveness, while allergens do not seem to provide relevant information. Based on these results, we conclude that Attention-based models can be used to assist ED managers in planning the ED staff. In addition, the Attention model is significantly smaller than the LSTM one. Therefore, its deployment requires less computation effort and it is more adequate to be used in lower-power computers, reducing maintenance costs.

Finally, it is important to note that ED admissions forecasting studies are highly dependent on the hospital context. Therefore, to adapt the models used here to other hospitals, specific studies would be necessary. The model configuration is crucial for achieving accurate results, as each hospital has its own characteristics and challenges. Also, it is important to identify relevant variables that should be included in the model. Only in this way, we can ensure that the models are suitable for application to different hospitals.

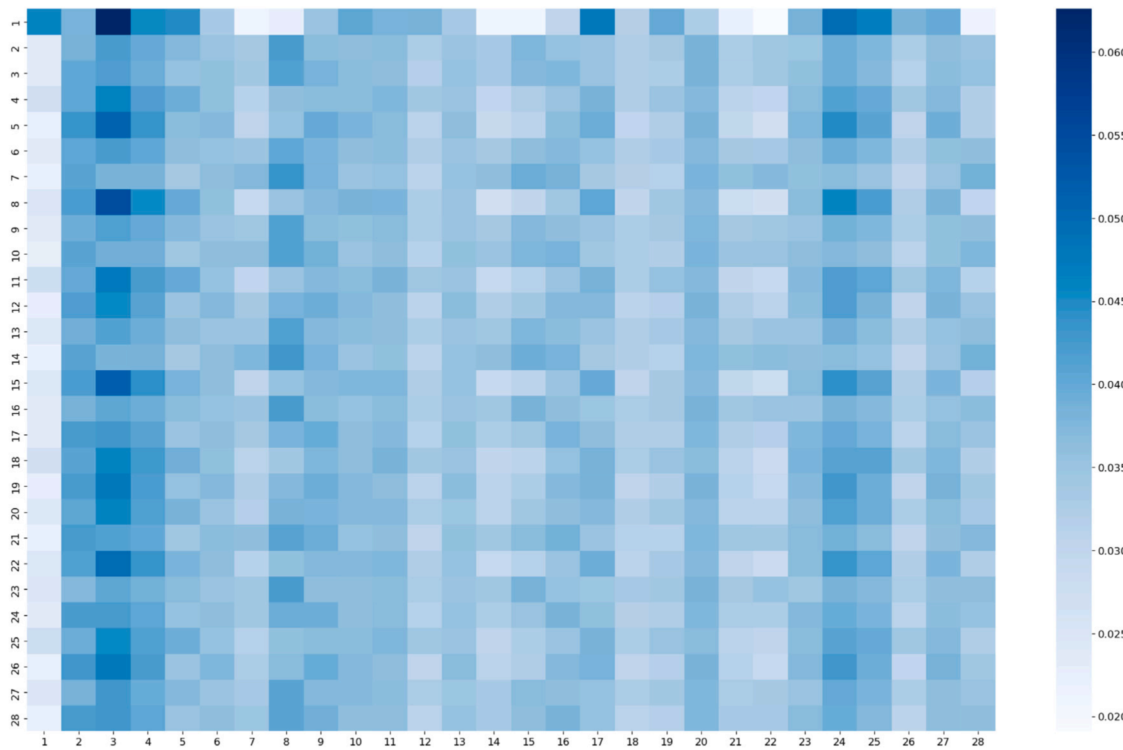

Fig. 18. Self attention patterns for Google Trends data.

CRediT authorship contribution statement

Hugo Álvarez-Chaves: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization. **Iván Maseda-Zurdo:** Data curation, Software, Methodology, Investigation, Writing – original draft, Visualization. **Pablo Muñoz:** Project administration, Conceptualization, Methodology, Investigation, Validation, Writing – original draft. **María D. R-Moreno:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This study was funded by Junta de Comunidades de Castilla-La Mancha, Spain (SBPLY/19/180501/000024). We would like to express our gratitude for the clinical support of Dr. Helena Hernández Martínez, Dr. María Isabel Pascual Benito and Mr. Francisco López Martínez. We also want to acknowledge the support of the Hospital Central de la Defensa Gómez Ulla (HCDGU), Madrid, Spain.

References

Almeida, H. S., Sousa, M., Mascarenhas, I., Russo, A., Barrento, M., Mendes, M., et al. (2022). The dynamics of patient visits to a public hospital pediatric emergency department: a time-series model. *Pediatric Emergency Care*, *38*(1), e240–e245. http://dx.doi.org/10.1097/PEC.0000000000002235.

Araz, O. M., Bentley, D., & Muelleman, R. L. (2014). Using Google Flu Trends data in forecasting influenza-like–illness related ED visits in Omaha, Nebraska. *The American Journal of Emergency Medicine*, *32*(9), 1016–1023. http://dx.doi.org/10.1016/j.ajem.2014.05.052.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. http://dx.doi.org/10.48550/arXiv.1409.0473, arXiv preprint arXiv:1409.0473.

Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., et al. (2012). Predicting emergency department admissions. *Emergency Medicine Journal*, *29*(5), 358–365. http://dx.doi.org/10.1136/emj.2010.103531.

Choudhury, A., & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. *British Journal of Healthcare Management*, *26*(1), 34–43. http://dx.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. http://dx.doi.org/10.1186/cc2981.

Diehl, A. K., Morris, M. D., & Mannis, S. A. (1981). Use of calendar and weather data to predict walk-in attendance. *Southern Medical Journal*, *74*(6), 709–712. http://dx.doi.org/10.1097/00007611-198106000-00020.

Duwalage, K. I., Burkett, E., White, G., Wong, A., & Thompson, M. H. (2020). Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emergency Medicine Australasia*, *32*(4), 618–625. http://dx.doi.org/10.1111/1742-6723.13481.

Ekström, A., Kurland, L., Farrokhnia, N., Castrén, M., & Nordberg, M. (2015). Forecasting emergency department visits using Internet data. *Annals of Emergency Medicine*, *65*(4), 436–442. http://dx.doi.org/10.1016/j.annemergmed.2014.10.008.

Fralick, M., Murray, J., & Mamdani, M. (2021). Predicting emergency department volumes: A multicenter prospective study. *The American Journal of Emergency Medicine*, *46*, 695–697. http://dx.doi.org/10.1016/j.ajem.2020.10.047.

Gul, M., & Celik, E. (2020). An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Systems*, *9*, 263–284. http://dx.doi.org/10.1080/20476965.2018.1547348.

Guttmann, A., Schull, M. J., Vermeulen, M. J., & Stukel, T. A. (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*, http://dx.doi.org/10.1136/bmj.d2983.

Healy, S., & Tyrrell, M. (2011). Stress in emergency departments: experiences of nurses and doctors. *Emergency Nurse*, *19*(4), http://dx.doi.org/10.7748/en2011.07.19.4.31.c8611.

Ho, A. F. W., To, B. Z. Y. S., Koh, J. M., & Cheong, K. H. (2019). Forecasting hospital emergency department patient volume using Internet search data. *IEEE Access*, *7*, 93387–93395. http://dx.doi.org/10.1109/ACCESS.2019.2928122.

Jiang, S., Liu, Q., & Ding, B. (2022). A systematic review of the modelling of patient arrivals in emergency departments. *Quantitative Imaging in Medicine and Surgery*, http://dx.doi.org/10.21037/QIMS-22-268.

Jilani, T., Housley, G., Figueredo, G., Tang, P.-S., Hatton, J., & Shaw, D. (2019). Short and Long Term predictions of hospital emergency department attendances. *International Journal of Medical Informatics*, *129*, 167–174. http://dx.doi.org/10.1016/j.ijmedinf.2019.05.011.

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

Kadri, F., Baraoui, M., & Nouaouri, I. (2019). An LSTM-based Deep Learning approach with application to predicting hospital emergency department admissions. In *2019 international conference on industrial engineering and systems management* (pp. 1–6). IEEE, http://dx.doi.org/10.1109/IESM45758.2019.8948130.

Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., et al. (2019). Time2Vec: Learning a vector representation of time. http://dx.doi.org/10.48550/arXiv.1907.05321.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980.

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. http://dx.doi.org/10.1111/acem.12182.

McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., & Kelen, G. D. (2008). The challenge of predicting demand for emergency department services. *Academic Emergency Medicine*, *15*(4), 337–346. http://dx.doi.org/10.1111/j.1553-2712.2008.00083.x.

Menke, N. B., Caputo, N., Fraser, R., Haber, J., Shields, C., & Menke, M. N. (2014). A retrospective analysis of the utility of an Artificial Neural Network to predict ED volume. *The American Journal of Emergency Medicine*, *32*(6), 614–617. http://dx.doi.org/10.1016/j.ajem.2014.03.011.

Petsis, S., Karamanou, A., Kalampokis, E., & Tarabanis, K. (2022). Forecasting and explaining emergency department visits in a public hospital. *Journal of Intelligent Information Systems*, *59*(2), 479–500. http://dx.doi.org/10.1007/s10844-022-00716-6.

Rema, V., & Sikdar, K. (2021). Time Series Modelling and Forecasting of Patient Arrivals at an Emergency Department of a Select Hospital. In *Recent trends in signal and image processing: ISSIP 2020* (pp. 53–65). Springer, http://dx.doi.org/10.1007/978-981-33-6966-5_6.

Rocha, C. N., & Rodrigues, F. (2021). Forecasting emergency department admissions. *Journal of Intelligent Information Systems*, 1–20. http://dx.doi.org/10.1007/s10844-021-00638-9.

Saygın, M., Gonca, T., Öztürk, Ö., Has, M., Çalışkan, S., Has, Z. G., et al. (2017). To investigate the effects of air pollution (PM10 and SO2) on the respiratory diseases asthma and chronic obstructive pulmonary disease. *Turkish Thoracic Journal*, *18*(2), 33. http://dx.doi.org/10.5152/turkthoracj.2017.16016.

Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., et al. (2017). Time to treatment and mortality during mandated emergency care for sepsis. *New England Journal of Medicine*, *376*(23), 2235–2244. http://dx.doi.org/10.1056/NEJMoa1703058.

Silva, E., Pereira, M. F., Vieira, J. T., Ferreira-Coimbra, J., Henriques, M., & Rodrigues, N. F. (2022). Predicting hospital emergency department visits accurately: A systematic review. *The International Journal of Health Planning and Management*, http://dx.doi.org/10.1002/hpm.3629.

Sudarshan, V. K., Brabrand, M., Range, T. M., & Wiil, U. K. (2021). Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. *Computers in Biology and Medicine*, *135*, Article 104541. http://dx.doi.org/10.1016/j.compbiomed.2021.104541.

Sun, Y., Heng, B. H., Seow, Y. T., & Seow, E. (2009). Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine*, *9*(1), 1–9. http://dx.doi.org/10.1186/1471-227X-9-1.

Suokas, J., & Lönnqvist, J. (1989). Work stress has negative effects on the attitudes of emergency personnel towards patients who attempt suicide. *Acta Psychiatrica Scandinavica*, *79*(5), 474–480. http://dx.doi.org/10.1111/j.1600-0447.1989.tb10290.x.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. *Operations Research for Health Care*, *21*, 1–18. http://dx.doi.org/10.1016/j.orhc.2019.01.002.

Xu, M., Wong, T. C., & Chin, K. S. (2013). Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using Artificial Neural Network. *Decision Support Systems*, *54*(3), 1488–1498. http://dx.doi.org/10.1016/j.dss.2012.12.019.

Xu, M., Wong, T. C., Chin, K. S., Wong, S., & Tsui, K. L. (2011). Modeling patient visits to accident and emergency department in Hong Kong. In *2011 IEEE international conference on industrial engineering and engineering management* (pp. 1730–1734). IEEE, http://dx.doi.org/10.1109/IEEM.2011.6118212.

Yousefi, M., Yousefi, M., Fathi, M., & Fogliatto, F. S. (2020). Patient visit forecasting in an emergency department using a Deep Neural Network approach. *Kybernetes*, *49*(9), 2335–2348. http://dx.doi.org/10.1108/K-10-2018-0520.

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A Transformer-based Framework for Multivariate Time Series Representation Learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 2114–2124). http://dx.doi.org/10.1145/3447548.3467401.

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