Received: 4 October 2021 | Revised: 28 September 2022 | Accepted: 4 February 2023

DOI: 10.1002/hpm.3629

REVIEW ARTICLE

WILEY

Predicting hospital emergency department visits accurately: A systematic review

Eduardo Silva[1] | Margarida F. Pereira[1] | Joana T. Vieira[2] | João Ferreira-Coimbra[2] | Mariana Henriques[3] | Nuno F. Rodrigues[4,5,6]

[1]University of Minho, Braga, Portugal

[2]University Hospital Center of São João, Porto, Portugal

[3]Centre of Biological Engineering, University of Minho, Braga, Portugal

[4]INESC TEC, Porto, Portugal

[5]Algoritmi Research Center, University of Minho, Braga, Portugal

[6]2Ai – School of Technology, IPCA, Barcelos, Portugal

Correspondence
Eduardo Silva.
Email: up202204602@eu.fe.up.pt

Abstract

Objectives: The emergency department (ED) is a very important healthcare entrance point, known for its challenging organisation and management due to demand unpredictability. An accurate forecast system of ED visits is crucial to the implementation of better management strategies that optimise resources utilization, reduce costs and improve public confidence. The aim of this review is to investigate the different factors that affect the ED visits forecasting outcomes, in particular the predictive variables and type of models applied.

Methods: A systematic search was conducted in PubMed, Web of Science and Scopus. The review methodology followed the PRISMA statement guidelines.

Results: Seven studies were selected, all exploring predictive models to forecast ED daily visits for general care. MAPE and RMAE were used to measure models' accuracy. All models displayed good accuracy, with errors below 10%.

Conclusions: Model selection and accuracy was found to be particularly sensitive to the ED dimension. While ARIMA-based and other linear models have good performance for short-time forecast, some machine learning methods proved to be more stable when forecasting multiple horizons. The inclusion of exogenous variables was found to be advantageous only in bigger EDs.

KEYWORDS
emergency department, forecasting, hospital, predictive models, resource management, visits

Highlights

- This work provides an insight into the study of forecasting emergency department (ED) visits, exploring different methodologies.
- Through the extracted data, it sets a unique analysis on how study features affect model's performance.
- It establishes a baseline knowledge for future studies that aim to explore these models.
- These forecasting models are important for hospital resource management planning.

1 | INTRODUCTION

Healthcare systems are a pillar of any modern society, playing an important role by directly answering the most urgent health needs of a population and building a feeling of security that greatly contributes to a general sentiment of well-being. The ED is a very important and often the first access point to the healthcare system. Its performance is paramount to the overall quality of the system and its public opinion.

The digital transformation is opening new opportunities to improve healthcare organisation and management, with digital data playing a central role. In the case of the ED, many of the opportunities for improvement rest upon data cycles capable of providing accurate predictions about future visits, such as the expected number of visits during different time periods, the distribution over different ED medical specialities, as well as forecast needs of human resources, equipment, facilities and medication.

Over the last decades, EDs have experienced an increasing demand in urban areas, putting a great amount of pressure over healthcare institutions that struggle to provide an adequate response.[1] ED's overcrowding is the most frequently reported problem by healthcare organisations worldwide.[2] The overwhelming and unpredictable increase of demand, particularly in certain periods of the year, leads to a mismatch between supply and demand, making it the prevailing reported cause.[3,4]

Most mitigation strategies proposed to overcome these problems include internal ED-specific and external hospital wide systemic adjustments, such as resource management, staff re-allocation and the smoothing of elective cases.[5] One of the most relevant contributions that data analytics can bring to the management strategy is to predict ED demand, which would enable hospitals to allocate, in advance, the different human and material resources and address the different potential deficiencies of services beforehand.[6,7] The development of models on emergency census indicates a need for predicting patient visits more accurately,[6,8] and these forecasting systems are capable of such.[6]

Forecasting models need previous data of any given time series to predict future behaviour,[9] and for ED demand, past ED visits and its daily, weekly and monthly variation are naturally the main data source for these models. The most prominent models are autoregressive integrated moving average (ARIMA),[10] linear regression (LR),[7,11,12] Poisson regression[13] and exponential smoothing (ES).[14,15] Machine learning models have also been recently used,[16] with the most frequent approaches resorting to different topologies of artificial neural networks (ANN).[17,18] On the other hand, some researchers consider that a single linear or nonlinear model may not be totally sufficient to capture all the characteristics of these time series, thus proposing hybrid or combined models to improve overall forecasting accuracy.[10]

Aside from calendar variables retrieved from the time-series data (day-of-the-week, holidays, weekends), some studies have considered other exogenous variables which are collected outside and tested for their ability to improve the model's performance. The most common sources of this type of data found in the literature include Internet search volumes for various query keywords,[19] hourly traffic flow in selected roads,[20] and more frequently, meteorological factors like ambient temperature and humidity.[21,22]

The objective of this systematic review is to analyse the panorama of forecasting ED visits, regarding the type of models and predictive variables applied, and parameters that may influence its performance. To the best of our knowledge, this is the first systematic review in which different models and variables are explored, with the last one, by Wargon et al.[6] in 2009, focussing only on mathematical techniques.

2 | MATERIALS AND METHODS

2.1 | Search strategy

The systematic review was conducted according to the PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration.[23] The literature selection and review process study were conducted to answer the following PICO question: 'What is the current state in forecasting ED visits by the general population, including models and incorporated variables, for a better overall predictive performance?'.

A literature research was carried by searching articles published after 2015 to May 2021 in Scopus, PubMed and Web of Science electronic databases. In all databases, the articles were retrieved by most recent, accessing all types of studies.

The search was carried using the online search engines of the databases using the keywords: forecasting, models, ED visits and resource management. To improve the accuracy of the search results, the following Boolean expression was built following the PICO question, using its keywords and synonyms: (forecasting AND ED AND (visits OR arrivals)) AND (method OR model) AND (better OR improved) AND (performance OR accuracy).

2.2 | Inclusion criteria

Screening inclusion criteria was: (1) articles must be written in English, (2) articles must focus on forecasting hospital visits, (3) articles need to focus on ED visits, (4) patient visits need to focus on urgent general care demand, without restrictions to any type of medical conditions.

2.3 | Study selection

Following the PICO strategy and the inclusion criteria, title and abstract for all articles were analysed by a single investigator. This stage allowed not only to include the articles that followed the PICO but also to exclude some that deviated from this review's main objective and the inclusion criteria. The remaining articles were assessed in full for a possible inclusion in the qualitative analysis.

2.4 | Research question

As a matter of coherence and conciseness, the initial PICO question was refined into eight more detailed questions that set a structured framework to guide the data extraction process from the selected articles. These questions assist the process of retrieving the unique features of each article in relation to the different subjects of interest, namely data characteristics, type of predictive models applied, variables and its outcomes. The first four questions regard the data sources from each article, the following two the models applied and the last two, enquiry the outcomes. These questions are:

RQ1: What was the volume of historical ED patient visits collected?
RQ2: In which geographical area are the ED patient visits inserted?
RQ3: What was the timespan of the data extracted?
RQ4: How many patients visit each ED per day?
RQ5: What types of models were used and to which other models were they compared to?
RQ6: Which exogenous predictive variables were considered in those models?
RQ7: What was the validation period used to calculate the model's parameters?
RQ8: What were the results obtained, especially each model's performance by evaluating its accuracy?

2.5 | Data extraction

The data extraction was performed independently by three investigators, who have closely followed the eight questions previously established, obtaining the following details from each article: a brief description of the study (forecasting horizon, type of forecasting); data characteristics from ED visits (volume, timespan, ED location); types of models that were applied and the exogenous predictive variables used to affect its performance (if used); and the outcomes, the best results for each model and MAPE (mean absolute percentage error) to measure model's accuracy. When an article also reports for models predicting patients by category, only the data and results for models predicting total patient visits were considered. Disagreements between answers to the questions were resolved by consensus between the three investigators after comparison and discussion of methods leading to each answer. The adopted solution was to highlight only the best model for each article.

3 | RESULTS

3.1 | Study selection

Initially, 223 articles were obtained from searching the electronic databases (161 from PubMed, 35 from Web of Science, 27 from Scopus) using the previously defined Boolean expression. After removing 31 repetitions (13,9%), 192 articles remained for further evaluation. Of these, 173 were removed in a preliminary title and/or abstract screening, by not complying with the review's inclusion criteria. From the remaining 19 potentially eligible reports, a more detailed examination was made accessing the full text. From these, 12 studies were excluded by one of the following reasons: (1) the objective was building a simulation model for modelling patient's visit,[24] (2) not providing a validation period out-of-train,[25] (3) focussing only on monthly or hourly forecasting rather than daily,[26–31] (4) models were only implemented in paediatric ED,[32–34] (5) approaching other accuracy measurements than MAPE, RMSE or RMAE.[35]

Some of the criteria adopted in this stage allowed to suppress some distinctions that would difficult their analysis. Choosing studies that predict daily ED visits, allows to focus on a single methodology and thus, excluding monthly and hourly predictions which are not the most suited for a short-term ED planning. Furthermore, methods that aim to predict hourly ED visits present high forecasting errors and need to be handled distinctively. Excluding paediatric EDs, also allowed to only focus on general care EDs from regular hospitals.

Finally, seven studies remained for further qualitative synthesis. The selection process is described in detail in the diagram of Figure 1.

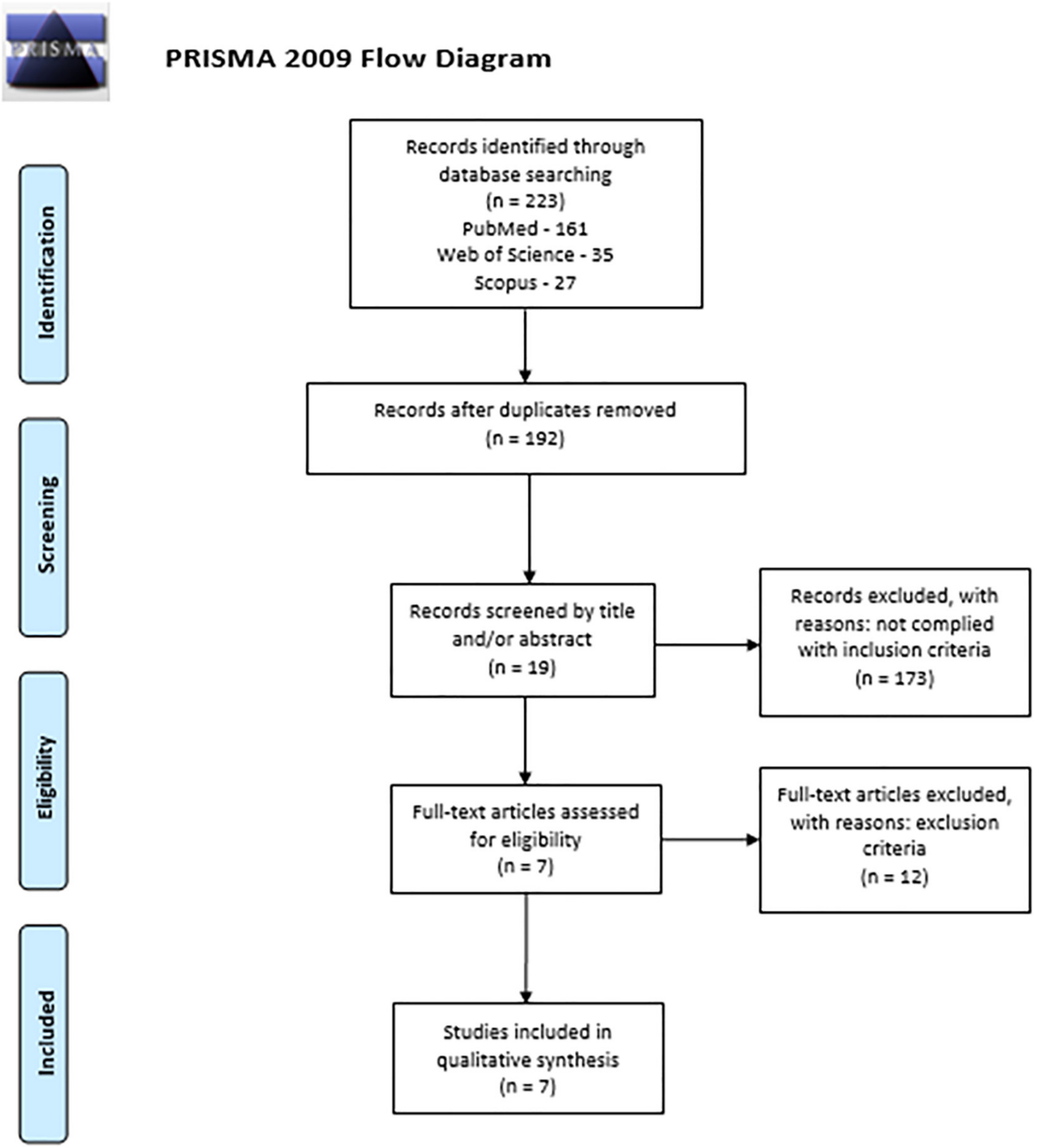

FIGURE 1 PRISMA flow diagram describing the study selection process.

3.2 | Summary of the results

In the seven studies included in this review, all forecasting models are constructed with daily time-series data, based on visits seeking healthcare for urgent treatment (Table S1). Concerning the amount of data used to train and test the models, the volume ranges from 57,128[22] to 252,438[36] ED daily visits, with three[37–39] out of seven studies not reporting this data. Studies' location varied, spreading across three continents and four different countries. All ED are from urban areas, though with very distinct dimensions, with the smallest having an average of 40 visits per day and the largest 700 visits per day, both coming from the study from Xu et al.[38] All studies applied the models to a single hospital, with the exception of two[38] that included multiple ED. The timespan of the time-series data varied from 1 year and 9 months[40,41] to 5 years[36,37] of records.

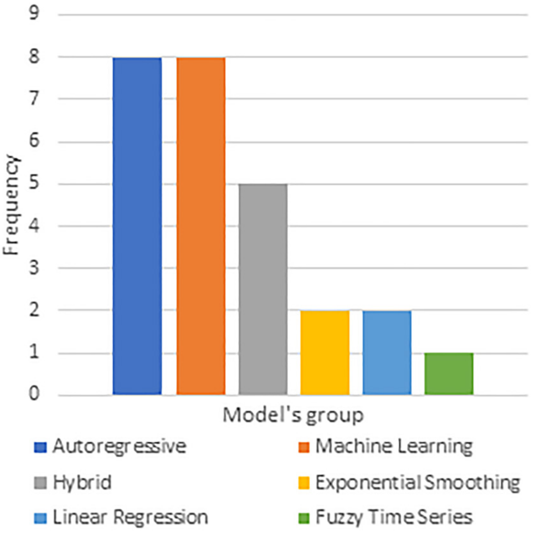

FIGURE 2 Frequency of the type of models used across the selected studies.

A large variety of different models was observed (Figure 2), with 20 different models being applied across the seven studies. However, most of them are derivations of the same model with additional features, which can lead one study to use multiple models from the same group. These models can be classified as (1) autoregressive moving average-based such as ARIMA, seasonal ARIMA (SARIMA), multiple seasonal ARIMA (MSARIMA), ARIMA with exogenous variables (ARIMAX); (2) machine learning, such as NN, deep neural network, long short-term memory, support vector machine, random forest; (3) hybrid models (ARIMA-LR, ARIMA-ANN, RNN-ARIMAX, RNN-SARIMA, RNN-PR); (4) ES-based such as seasonal multiplicative Holt-Winters, seasonal ES (SS); (5) LR-based such as generalised linear model (GLM); and (6) fuzzy time series (FTS). The most frequent are ARIMA-based, and machine learning used 8 times each in the selected articles.

In total, four studies[22,38,40,41] include exogenous predictive variables in the models, all related to meteorological factors (e.g., temperature, weather conditions).

Most studies use one or more parameters for measuring the model's accuracy. MAPE is the most reliable parameter for a direct comparison between different studies, as it is scale-independent, can be used for comparing forecasts on different scales. Still, as MAPE is dependent on the expected and forecasted values, it is also dependent of the validation period (the length for which it is calculated). All studies mention the period of the validation, which ranged from 28 days[40,41] to 589 days.[39] For most studies MAPE was obtained in each article, with exception for Afilal et al.,[36] that opted to use RMAE (relative mean absolute error) to measure its models' accuracy. The most accurate models obtained a MAPE ranging from 2.63%[37] to 12,79%,[41] each determined under very different circumstances. In a broader perspective, most studies obtained good outcomes, with the large majority achieving a MAPE around 5%.

4 | DISCUSSION

4.1 | Data

In time-series analysis, seasonality may emerge from any repetitive pattern that occurs periodically such as day-of-week, month-of-year, weekends, or weekly in ED data visits. This distinctive characteristic is a factor that can hold a strong predictive value if present in the data. Most studies found day-of-the-week as the most important pattern in time-series data, given that it generally produced better predictors. Day-of-the-week seasonality is complemented in all studies with data from more than 1 year of visits records to capture relevant seasonal patterns that occur in a year timespan. Month-of-year was also found to be a relevant pattern in some studies, with Jiang et al.[40] notably showing, through a genetic algorithm for feature selection, that this is a key feature affecting patient flow. On the contrary, Calegari et al.[22] mentioned that month-of-year has no effect on demand, showing not to be significant through an one-way ANOVA test. Holidays also hold contradictory results, with some studies[22,40] suggesting that these variables may not be relevant, while Yousefi et al.[39] presented an univariate analysis to support its statistical significance.

Our analysis indicates that the reason for these conflicting results lays not on the different mathematical analysis processes taken by each study, but rather on substantial differences between the ED realities observed. One relevant factor contributing to these apparently contradictory results is the size of the ED. Xu et al.[38] verified that when forecasting visits in the department with the smallest volume of daily visits (ED B), seasonal patterns have not the same impact as in the other department (ED A), which has a larger number of visits. The importance of the ED size in forecasting visits is also confirmed by the opposite relevance of the month-of-year variable between Calegari et al.[22] and Jiang et al.,[40] where the former analysed an ED with only 48 attendances per day on average and the later an ED with 384 visits. Moreover, the data from the analysed studies points to a threshold of around 400 ED visits per day or more[39,40] above which calendar variables start gaining importance to improve the forecasting results of models.[22]

In sum, day-of-the-week has found to be the most relevant seasonal predictor, but month-the-year can still be a good predictor in certain situations, such as when the EDs have a high volume of visits. These specific characteristics of the ED and the data available plays a central role in the selection of forecasting models and the variables to take into consideration.

4.2 | Models and variables

ARIMA models are the most frequently used in the reviewed works. When dealing with a time series prediction problem, in which the features of randomness and cyclic fluctuations have the largest effect on forecasting accuracy, these models fit particularly well, extracting the linear features of randomness and trend.[10] Moreover, ARIMA follow a univariate approach method that require no external data.

Models such as SARIMA and MSARIMA are extensions to the basic ARIMA model, adding new features. SARIMA allows the incorporation of seasonal patterns and Calegari et al.[22] showed there is a strong autocorrelation at 7-day lag, captured when tunning the model's parameters. This suggests there is an evident seasonality which is directly associated with weekday patterns discussed above. MSARIMA incorporates independent explanatory variables to SARIMA in search for a better characterisation of the demand time series and was used[22] to include calendrical variables (day-of-the-week, month-of-year) and 8 meteorological factors (in a 1- to 7-day lag). By the nature of the data, where ED visits are heavily influenced by repeated population behaviours, this seasonality is to be expected. Once a persistent strong autocorrelation is found, SARIMA might be more suitable than other autoregressive models. ARMA is another autoregressive model experimented,[36] but despite achieving good results, with a RMAE of 8,16%, it only worked for stationary data, which may be a crucial drawback for data with obvious seasonal trends. However, ARIMA-based models can be complex in practice when it comes to choosing its parameters by autocorrelation plots and other not so precise methodologies. On the other hand, FTS does not require any seasonality and periodicity adjustments.

ARIMA models are very good when dealing with the linear parts, but weak when working with nonlinear data.[25] Since most complex time-series contain nonlinear components, machine learning methods are more capable of modelling the input-output relationship in complex and dynamic systems that contain these components, which contributes for being designated as non-linear methods.[42] Patient visits are very influenced by seasonal patterns and capturing this information is a key factor for achieving better forecasts. Linear models are more suitable if the data is heavily driven by seasonal patterns, while machine learning models are more appropriate for data that contains more randomness. Even though it needs some knowledge to manipulate its parameters, the class of ARIMA models are more intuitive to use, as they are adapted to deal with time-series data. On the other hand, machine learning models can only work with seasonal patterns as input variables, which involves a greater knowledge that hardens its interpretation. Models such as SARIMA, aside from being more computationally efficient, can capture relevant patterns without the need to introduce other algorithms like the feature selection applied by Jiang et al.[40]

Exogenous variables might be a solution to the unpredictability of some patterns, such as seasonality or trend changes, that might find explanatory meaning with these predictors. Following this line of enquiry, Calegari et al.[22] showed that maximum and minimal temperatures, with 0- and 1-day lag, are statistically significant when tested for independent variables. Jiang et al.[40] reached similar conclusions, while having a different approach by using a feature selection algorithm. In this case, variables such as absolute daily mean air temperature, mean dew point, mean relative humidity, and mean wind speed, were considered important. An important hypothesis that hasn't been explored by these studies is to find how the correlation to visits data can explain some random events that occur in portions of the data, sudden changes in trend or seasonality.

4.3 | Outcomes

The wide variety of models applied in each study leads to various outcomes, making it difficult to draw conclusions between the type of model and its accuracy. However, it is possible to gather a few noteworthy insights about model's adequacy according to some general characteristics of ED.

It is clear from the analysed works that hospitals with different sizes of ED visits need different approaches to obtain better accuracy, which suggests that model selection should also vary according to ED size. Xu et al.[38] using a ARIMA-LR model, found that smaller ED, with 40 visits per day, obtained a MAPE of 12,3%, in contrast with another ED which achieved a MAPE of 6,5%. Similar differences were observed for every model applied, which one believes to steam from how trends and seasonality impact the forecasting results. Patient flows in smaller hospitals are more randomised, making it difficult to find a relatively consistent seasonal pattern. On the contrary, bigger hospitals, that receive much more visits, experience more pronounced seasonal variations, which are easily detectable.[38]

However, the same conclusion cannot be asserted for the FTS model used by Jilani et al.[37] which obtained similar outcomes when applied to EDs with very different daily mean visits. In this case, the researchers implemented a weekday time series model which achieved a MAPE of 3,19%, 3,35%, 4,14% and 3,95%, in EDs with 212, 177, 92 and 64 daily mean visits, respectively. Even though there are differences in accuracy between the EDs with highest and lowest number of visits, these are not so significant when compared with the ones obtained by Xu et al.[38]

Looking at the outcomes of a SARIMA applied to two very different EDs located in different countries, Calegari et al.[22] and Jiang et al.[40] obtained a MAPE of 6,23% and 7,8%, respectively. This contrasts with Xu et al.[38] conclusion, that an ARIMA model is incapable of producing similar accuracy results under such ED size differences. The best explanation for these differences in ARIMA's performance may lay on the parametrisation used by Calegari et al.[22] which by adding a seasonal component to ARIMA (SARIMA) made the model more capable of capturing seasonal patterns of each particular ED. However, these findings must be taken with precaution, because the EDs of Calegari et al.[22] are from distinct regions, which can entail many different factors that affect these components.

Model accuracy is also heavily influenced by the forecast horizon. Calegari et al.[22] explored a wide range of models (Figure 3) with different forecasting horizons, all of which presenting decreased accuracy for longer horizons. Noteworthily, the largest decrease occurs when extending the forecasting horizon from 1-day to 7-days in advance, with MAPE rising from 2,91% to 10,67% using a SS model. Above 7-days, MAPE stabilises and only rises to 11,35% for 21-days in advance and to 11,51% for 28-days in advance. The same goes for every other model. Typically, traditional statistical models, like ARIMA or LR, work well for capturing the short memory and report poor accuracy when the prediction horizon is more than 1 week, making them unfit for long-term decision making.[40] Xu et al.[38] observed a similar occurrence when forecasting for 1- and 7-days in advance (Figure 3), with all models reporting a slightly

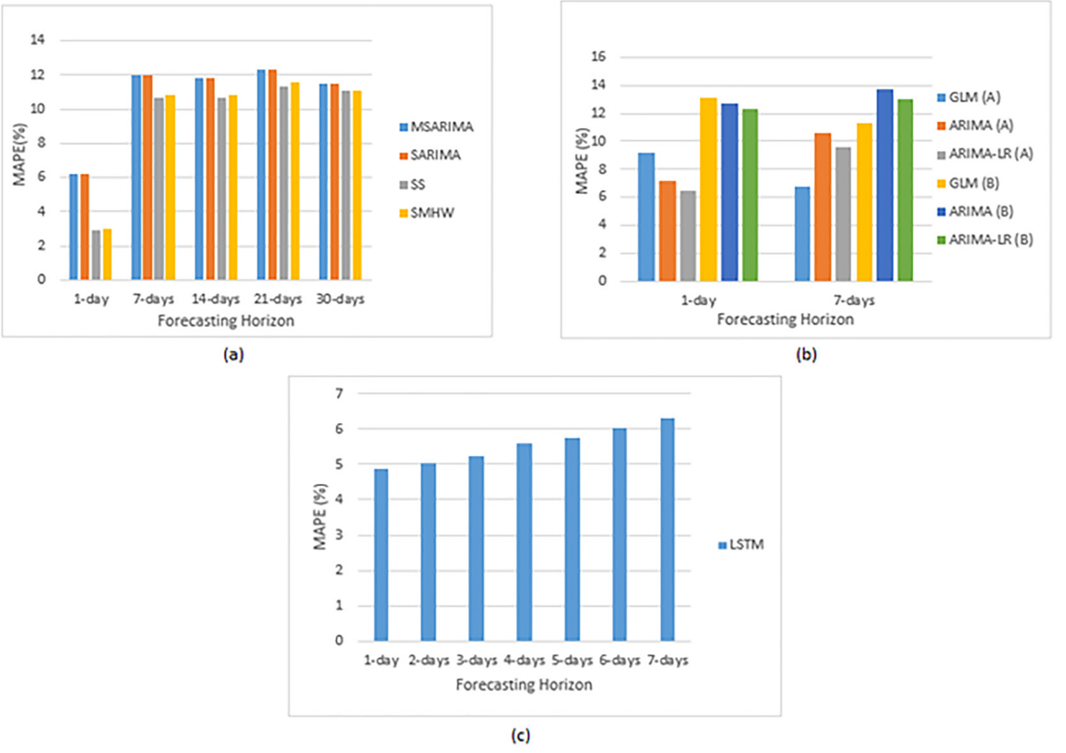

FIGURE 3 Forecasting accuracy by MAPE (%) under different horizons: (A) Calegari et al.[20], (B) Xu et al.[37] (A-applied in emergency department (ED) A; B-applied in ED B), (C) Yousefi et al.[38].

increase in MAPE for both EDs, even though more markedly in ED A (from 6,1% to 9,6%, with ARIMA-LR). Both hybrid models, ARIMA-LR and ARIMA-ANN, are the most accurate for 1- day-ahead forecasting, while for 7-days in advance, adaptive GLM shows the lowest MAPE. Generalised linear model also stands out for being the only model to have a better accuracy for a longer horizon, which opens again the possibility of having models that are more suitable for longer prediction horizons and others for short horizon. Long short-term memory also showed[39] good results when predicting up to 7-days ahead (Figure 3), with MAPE only increasing from 4,89% to 6,31%. On balance, linear models usually display good performance for short-term horizons, while machine learning (non-linear) are capable of achieving good accuracy for short and long forecasting horizons.[39–41]

In what respects to the usage of these forecasting results for healthcare management, namely for resource allocation and staff planning, forecasting up to 7-days is more important than a more long-term prediction horizon. This variation from 1-day to 7-day horizon, and the way it affects model's accuracy, explains the decision of using different models for different horizons.

Looking at the overall accuracy achieved by each type of model, it is not possible to assure the superiority of any of the models. Most models displayed good accuracy, achieving most of the times a MAPE under 10%. Only under some specific circunstances[22,38] do these underperform, such as for small EDs or when pushing the prediction horizon over 21 days. Fuzzy time series is of special mention here, presenting very promising results for longer forecasting horizons in multiple EDs, with a MAPE between 2.6% and 4.7% for a 4 week ahead forecasting. In this case, the authors took a different approach, developing models for each weekday, a strategy that certainly deserves further research to confirm the results for other ED realities.

Some articles explored a range of hybrid models with a good overall performance. Jiang et al.[41] used several RNN hybrid models capable of reaching a MAPE of 3,38% on a 28-days horizon. The attempt to use hybrid models was a response to try to combine the best individual advantages of two or more forecasting approaches in a single solution. Xu et al.,[38] were capable of this by reaching the best results with an ARIMA-LR and ARIMA-ANN model, excelling ARIMA and GLM models performance. But as mentioned above this is not the case for all forecasting horizons, failing to achieve the best results for a 7-day horizon forecast, which shows once again how the choice of the type of model is always dependant on a series of factors.

Validation period is also an important factor affecting these forecasts. Generally, a bigger period provides a better validation, as the models operate over a wider timespan which increases the probability of capturing relevant calendar events. To this extent, it is important to understand how the models react to certain periods of the year and how the accuracy evolves further in time. A small validation period may deceive the interpretation of the forecast if, for example, that period is typically low on its predictive accuracy. Even though some studies provide a large validation period, this strategy was only followed by one work, which may explain some of discrepancies found between results. Jiang et al.[40] were the only to provide a segmented validation analysis, over six periods of 28 days. DNN-I-GA, the most accurate model, achieved a MAPE ranging from 3,97% to 8,12%, which shows how volatile and dependent these can be on the chosen period.

In what respects to exogenous variables, Calegari et al.[22] found that using MSARIMA models considering meteorological factors did not improve the performance of the SARIMA models, as the MAPE remained at 6.23%, although they showed to be statistically relevant before. Jiang et al.,[40] determined that GLM and ARIMAX (including meteorological factors) perform better than ANN and ARIMA. From the point of view of data, the former indicate that patient arrival presents a more linear relationship with calendar-based variables and meteorological factors, which cannot only be explained by historical trend alone. Corroborating these results, Xu et al.[38] concludes that including external variables improves forecasting performance. However, this only applies to the ED with high patient flow, which implies that for small departments', visits are probably driven by serial correlations instead of exogenous indicators.

Several reasons may contribute, possibly in conjunction, to these contradicting results. Firstly, the association of such variables is dependent on cultural population habits in each region, which may result in significant differences on data correlation. Then, the dimension of the ED seems to have an important role, as the studies that reported poor performance from these variables are all related to small EDs, with only 68[22] and 40[38] daily mean visits. As these have few visits, data tends to be more randomized, and seasonality and other patterns are not as detectable as in larger EDs. When trying to establish a relation between external predictors and ED visits data, the models will certainly fail as the data patterns do not reflect the general population behaviour.

While exogenous variables seem to hold great potential, they have been largely unexplored in the literature, where one cannot find any systematic evaluation of their contribution to the prediction's accuracy, nor direct comparison between different variables. There is a lack of understanding about their significance under different context circumstances, such as regions, hospitals sizes and healthcare systems organisation.

4.4 | Limitations

The reviewed articles had several limitations that may hamper some of the conclusions drawn in this systematic review. First, the studies are carried out in several different countries, covering health systems with different characteristics, which makes all the assumptions about the model's performance and variables dependent on this factor. The validation period used may also deceive our interpretation of the accuracy measured, as it changes according to the period of the year for which it is being calculated. Publication bias could also contribute to a deceiving view on some of these findings, for example, poorly performing models are less likely to be published, yet may contain valuable information about variables that should or should not be used.

The review process also includes limitations related to the searching method. Even though different search queries were tested and results compared with the final query for any relevant missing articles, some relevant studies could still have been excluded. Bias could have been induced by restricting criteria to only English-language publications. Despite some studies having included models to forecast different categorisations of patients, those details are beyond the purpose of this review and were not considered.

Other limitations are related to the impossibility to retrieve data from a few selected studies. Three articles[37–39] did not provide the total volume of ED visits included in the models, and other[36] included a different parameter to measure the accuracy of its models.

4.5 | Recommendation for future studies

After the detailed analysis carried out in this review, we summarise the most relevant insights for future studies on the subject as follows:

- Use at least 3 years of data, to allow to validate the models through a segmented period and study how the models perform throughout the year.
- ARIMA-based models are a great option if the data is heavily driven by seasonal patterns, with calendar variables being statistically significant, (usually for larger ED), while machine learning methods for more randomized data (usually smaller ED).
- If there is a strong autocorrelation in the data, use SARIMA over other autoregressive models, as it allows to incorporate seasonal patterns.
- Use different models for short and longer forecasting horizons, usually for 1 to 3-day and 4 to 7-day horizon, respectively.
- Only use machine learning methods when forecasting for a horizon over a 1-day or for different horizons.
- Hybrid models may be a great alternative for forecasts with 1-day horizon, as they combine the best of each model.
- Meteorological factors are an unreliable exogenous predictor that depend on other different factors, such as the size or the location of the ED.
- Future studies presenting new models should present important results data to compare them with other studies and better contribute to the improvement of ED visits forecast. To this extent, studies should present hospital size, total number of ED visits, and results with and without exogenous variables.

5 | CONCLUSIONS

The insights we provide about forecasting ED visits draw the general conclusion that each specific context will need different approaches, depending more on factors such as forecast horizon and ED dimension, and less on the type of model chosen.

Predicting visits for EDs which have low daily patient flow needs innovative approaches, as distinct factors have shown to yield unsatisfactory predictors in this case, contradicting the results obtained for larger EDs. In these, patient flow historical data and calendar variables are often enough to provide an excellent prediction accuracy. Finding new variables that can correlate with visits' data and become strong and reliable predictors is an area where much research is still needed.

The ability to predict patients' flow days in advance may be a simple but challenging way of providing an important contribution to solve several problems of ED management. A perspective that surpasses the analytic insights provided by the statistical analysis of ED visits is displayed here, opening new perspectives on the subject, and clearing a new path for future studies.

AUTHOR CONTRIBUTIONS

Eduardo Silva - acquisition of the data, analysis and interpretation of the data, drafting of the manuscrip; Margarida F. Pereira, Joana T. Vieira - analysis and interpretation of the data; João Ferreira-Coimbra, Mariana Henriques, Nuno F. Rodrigues - critical revision of the manuscript for important intellectual content.

ACKNOWLEDGEMENTS

This work was carried out during Eduardo's master thesis, at University of Minho, as part of a longer investigation on the forecast of visits to the emergency department.

CONFLICTS OF INTEREST STATEMENT

Eduardo Silva reports no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICS STATEMENT

Not applicable.

ORCID

Eduardo Silva https://orcid.org/0000-0002-5433-8301
Joana T. Vieira https://orcid.org/0000-0003-3110-1528

REFERENCES

1. Kellermann AL. Crisis in the emergency department. *N Engl J Med*. 2006;355(13):1300-1303. https://doi.org/10.1056/nejmp068194
2. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions; 2018.
3. Di Somma S, Paladino L, Vaughan L, Lalle I, Magrini L, Magnanti M. Overcrowding in emergency department: an international issue. *Intern Emerg Med*. 2015;10(2):171-175. https://doi.org/10.1007/s11739-014-1154-8
4. Lindner G, Woitok BK. Emergency department overcrowding: analysis and strategies to manage an international phenomenon. *Wien Klin Wochenschr*. 2020;133(5-6):229-233. https://doi.org/10.1007/s00508-019-01596-7
5. Salway R, Valenzuela R, Shoenberger J, Mallon W, Viccellio A. Emergency department (Ed) overcrowding: evidence-based answers to frequently asked questions. *Rev Médica Clínica Las Condes*. 2017;28(2):213-219. https://doi.org/10.1016/j.rmclc.2017.04.008
6. Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. *Emerg Med J*. 2009;26(6):395–399. https://doi.org/10.1136/emj.2008.062380
7. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170. https://doi.org/10.1111/j.1553-2712.2007.00032.x
8. Flottemesch TJ, Gordon BD, Jones SS. Advanced statistics: developing a formal model of emergency department census and defining operational efficiency. *Acad Emerg Med*. 2007;14(9):799–809. https://doi.org/10.1197/j.aem.2007.05.011
9. Chatfield C. *Time-Series Forecasting*. 1st ed.: Chapman and Hall/CRC; 2000. https://www.taylorfrancis.com/books/mono/10.1201/9781420036206/time-series-forecasting-chris-chatfield
10. Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Serv Res*. 2017;17(1):1–13, https://doi.org/10.1186/s12913-017-2407-9
11. Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med*. 2010;17(9):970–978. https://doi.org/10.1111/j.1553-2712.2010.00847.x
12. Hertzum M. Forecasting hourly patient visits in the emergency department to counteract crowding. *Ergon Open J*. 2017;10(1):1-13. https://doi.org/10.2174/1875934301710010001
13. McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15(4):337–346. http://doi.wiley.com/10.1111/j.1553-2712.2008.00083.x
14. Carvalho-Silva M, Monteiro MTT, de Sá-Soares F, Dória-Nóbrega S. Assessment of forecasting models for patients arrival at Emergency Department. *Oper Res Heal Care*. 2018;18:112-118. https://doi.org/10.1016/j.orhc.2017.05.001
15. Bergs J, Heerinckx P, Verelst S. Knowing what to expect, forecasting monthly emergency department visits: a time-series analysis. *Int Emerg Nurs*. 2014;22(2):112-115. https://doi.org/10.1016/j.ienj.2013.08.001
16. Tealab A. Time series forecasting using artificial neural networks methodologies: a systematic review. *Futur Comput Informatics J*. 2018;3(2):334-340. https://doi.org/10.1016/j.fcij.2018.10.003
17. Gul M, Guneri AF. Planning the future of emergency departments: forecasting ed patient arrivals by using regression and neural network models. *Int J Ind Eng Theory Appl Pract*. 2016;23(2):137-154.

18. Menke NB, Caputo N, Fraser R, Haber J, Shields C, Menke MN. A retrospective analysis of the utility of an artificial neural network to predict ED volume. *Am J Emerg Med*. 2014;32(6):614-617. https://doi.org/10.1016/j.ajem.2014.03.011
19. Ho AFW, To BZYS, Koh JM, Cheong KH. Forecasting hospital emergency department patient volume using internet search data. *IEEE Access*. 2019;7:93387-93395. https://doi.org/10.1109/access.2019.2928122
20. Rauch J, Hübner U, Denter M, Babitsch B. Improving the prediction of emergency department crowding: a time series analysis including road traffic flow. *Stud Health Technol Inform*. 2019;260:57-64.
21. Marcilio I, Hajat S, Gouveia N. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Acad Emerg Med*. 2013;20(8):769-777. https://doi.org/10.1111/acem.12182
22. Calegari R, Fogliatto FS, Lucini FR, Neyeloff J, Kuchenbecker RS, Schaan BD. Forecasting daily volume and acuity of patients in the emergency department. *Comput Math Methods Med*. 2016:1-8. https://doi.org/10.1155/2016/3863268
23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339. b2700. https://doi.org/10.1136/bmj.b2700
24. Ordu M, Demir E, Tofallis C. A decision support system for demand and capacity modelling of an accident and emergency department. *Heal Syst*. 2020;9(1):31–56. https://www.tandfonline.com/doi/abs/10.1080/20476965.2018.1561161
25. Yucesan M, Gul M, Celik E. A multi-method patient arrival forecasting outline for hospital emergency departments. *Int J Healthc Manag*. 2018;0(0):1-13. https://doi.org/10.1080/20479700.2018.1531608
26. Choudhury A, Urena E. Forecasting hourly emergency department arrival using time series analysis. *Br J Heal Care Manag*. 2020;26(1):34–43. https://doi.org/10.12968/bjhc.2019.0067
27. Aboagye-Sarfo P, Mai Q, Sanfilippo FM, Preen DB, Stewart LM, Fatovich DM. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *J Biomed Inform*. 2015;57:62–73. https://doi.org/10.1016/j.jbi.2015.06.022
28. Juang WC, Huang SJ, Huang FD, Cheng PW, Wann SR. Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ*. 2017;7(11). e018628, https://doi.org/10.1136/bmjopen-2017-018628
29. Duarte D, Walshaw C, Ramesh N. A comparison of time-series predictions for healthcare emergency department indicators and the impact of COVID-19. *Appl Sci*. 2021;11(8):3561. https://doi.org/10.3390/app11083561
30. Cheng Q, Argon NT, Evans CS, Liu Y, Platts-Mills TF, Ziya S. Forecasting emergency department hourly occupancy using time series analysis. *Am J Emerg Med*. 2021;48:177–82. https://pubmed.ncbi.nlm.nih.gov/33964692/
31. Asheim A, Bache-Wiig Bjørnsen LP, Næss-Pleym LE, Uleberg O, Dale J, Nilsen SM. Real-time forecasting of emergency department arrivals using prehospital data. *BMC Emerg Med*. 2019;19(1):1–6. https://doi.org/10.1186/s12873-019-0256-z
32. Almeida HS, Sousa M, Mascarenhas I, et al. The dynamics of patient visits to a public hospital pediatric emergency department. *Pediatr Emerg Care*. 2020. Publish Ah.
33. Tideman S, Santillana M, Bickel J, Reis B. Internet search query data improve forecasts of daily emergency department volume. *J Am Med Informatics Assoc*. 2019;26(12):1574–1583. https://doi.org/10.1093/jamia/ocz154
34. Harrou F, Dairi A, Kadri F, Sun Y. Forecasting emergency department overcrowding: a deep learning framework. 2020.
35. Pekel E, Gul M, Celik E. Forecasting daily patient visits in an emergency department by ga-ann hybrid approach. *Proc 14th Int Symp Oper Res SOR*. 2017:473-478.
36. Afilal M, Yalaoui F, Dugardin F, Amodeo L, Laplanche D, Blua P. Forecasting the emergency department patients flow. *J Med Syst*. 2016;40(7):175. https://doi.org/10.1007/s10916-016-0527-0
37. Jilani T, Housley G, Figueredo G, Tang PS, Hatton J, Shaw D. Short and Long term predictions of Hospital emergency department attendances. *Int J Med Inform*. 2019;129:167-174. https://doi.org/10.1016/j.ijmedinf.2019.05.011
38. Xu Q, Tsui KL, Jiang W, Guo H. A hybrid approach for forecasting patient visits in emergency department. *Qual Reliab Eng Int*. 2016;32(8):2751-2759. https://doi.org/10.1002/qre.2095
39. Yousefi M, Yousefi M, Fathi M, Fogliatto FS. Patient visit forecasting in an emergency department using a deep neural network approach. *Kybernetes*. 2020;49(9):2335-2348. https://doi.org/10.1108/k-10-2018-0520
40. Jiang S, Chin KS, Tsui KL. A universal deep learning approach for modeling the flow of patients under different severities. *Comput Methods Programs Biomed*. 2018;154:191-203. https://doi.org/10.1016/j.cmpb.2017.11.003
41. Jiang S, Xiao R, Wang L, et al. Combining deep neural networks and classical time series regression models for forecasting patient flows in Hong Kong. *IEEE Access*. 2019;7:118965-118974. https://doi.org/10.1109/access.2019.2936550
42. Bao W. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One* 2017;12(7). e0180944, https://doi.org/10.1371/journal.pone.0180944

AUTHOR BIOGRAPHIES

Eduardo Silva graduated in Biomedical Engineering, and he is currently a PhD researcher in Digital Media at University of Porto. His research Allies the areas of human-computer interaction, clinical engineering, and digital medicine. In his master thesis he has worked with forecasting models for predicting emergency department visits.

Margarida F. Pereira obtained her MSc in Biomedical Engineering at the University of Minho, Portugal. She has focussed her studies in the developed of a preliminary serious immersive Virtual Reality game based on hand tracking to complement conventional occupational therapy in hand rehabilitation.

Joana T. Vieira obtained her MD in Medicine at University of Porto. She's currently an Internal Medicine resident in Centro Hospitalar Universitário São João, Porto. Her research interests are acute care, internal medicine, and cardiovascular diseases.

João Ferreira-Coimbrao is an Internal Medicine Specialist, currently Felipe in intensive care medicine at the Intensive Care Medicine Department in Centro Hospitalar Universitário São João, Porto. His research interests are the acute care, ICU epidemiology, low respiratory tract infections, severe respiratory infections, as well as human-computer interaction and digital medicine.

Mariana Henriques is an Assistant Professor at the Department of Biological Engineering at University of Minho and a researcher at the Centre for Biological Engineering, in the Biofilms group, focussed on microbiology applied to medicine. She was graduated in Biological Engineering and has a PhD in Chemical and Biological Engineering, both at University of Minho.

Nuno F. Rodrigues is an Associate Professor of Science and Technology of Programing at the School of Technology of the Polytechnic Institute of Cávado e Ave. He graduated in Mathematics and Computer Science and obtained a PhD in Computer Engineering. He has coordinated and participated in several research projects in different application areas of computer science and digital technologies, having published several articles.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Silva E, Pereira MF, Vieira JT, Ferreira-Coimbra J, Henriques M, Rodrigues NF. Predicting hospital emergency department visits accurately: a systematic review. *Int J Health Plann Mgmt*. 2023;1-14. https://doi.org/10.1002/hpm.3629