Developing predictive analytic tools to forecast the influx of emergency patients

Diana Silva Serrano

diana.serrano@tecnico.ulisboa.pt

Thesis to obtain the Master of Science Degree in Biomedical Engineering

Instituto Superior Técnico, Universidade de Lisboa

June 2020

Abstract

The Emergency Department (ED) demand is the key factor in managing resources in any ED. Therefore, the objective of this work is focused on the development of an informative ED demand predictive tool. This information can assist in the allocation of resources and minimizing health hazards associated with long waiting times. The literature presents several methodologies for similar problems; however, none of them considers the knowledge of health professionals. A hybrid methodology exhibits better chances in obtaining high accuracy for forecasting problems. Therefore, a methodology that combines the expertise of professionals with mathematical analysis of historical data was developed and tested in the *Hospital Lusíadas Lisboa* context, where professionals identified external factors, such as temperature, that impact the ED demand. A combination of methods that examine seasonality, such as the Autorregressive Integrated Moving Average, were trained and tested using supervised learning that allows the identification of the models that produce the best forecasts. These models use historical data and, in some cases, external factors. A Decision Support System (DSS), named ForecastER ®, was developed to integrate the predicting tool, facilitating its use by health professionals. In collaboration with them, the inclusion criteria for the DSS were established. In a real-life situation, the DSS exhibits inferior accuracy when compared to the modelling phase since part of the test took place during a pandemic. Finally, this system can be used in other EDs, exhibiting the potential to be integrated into the ED management system and the possibility to be applied in real-time.

1. Introduction

The health system and the Emergency Department (ED) hold a prominent and essential role in any society, as well as having the main share of human, financial, and equipment resources usage. The ED is specialized in emergency medicine, being responsible for the first medical care to critical patients. It does not require an appointment and therefore, it has to be able to provide the necessary treatment in all medical specialities without previous requests. In Portugal, between 2012 and 2018, the total number of ED visits increased by 10.8%. In the same period, the number of ED visits decreased by 7.4% in public hospitals; however, in the private sector, visits raised 63.2% [1].

The Manchester Triage System (MTS) is applied, in Portuguese EDs, to prioritize the need for health care [2]. Thus, there are different emergency categories, such as white, blue or green for less urgent cases and yellow, orange and red in more serious health problems [2]. It can be easily gathered from the National Health System (NHS) data that it is normal to have non-emergency episodes in the ED [3]. These patients consume resources and contribute to a major problem, termed overcrowding. The ED is a critical department in a hospital, that can not fail yet, it can not consume infinite resources. Consequently, hospitals should be able to predict the necessary resources for their ED. These are directly linked to ED demand. Therefore, if a hospital forecasts the ED demand, the same can create an adequate organizational plan that would satisfy the demand of the department without wasting resources. Nevertheless, this process is not a simple task since the primary difficulty is obtaining a trustworthy forecast linked to the volatility of ED visit.

The correct prediction of a future event or series of events is denominated forecast. Forecasting correctly can be a challenging task, but it is extremely important in many different fields, such as economy, medicine, or government. It is a process that uses mathematical models capable of making future predictions by assuming that one or more variables (e.g. time) affect the event to be forecast. There are different techniques, used in forecasting, that can be divided into two restrictive categories, for instance, qualitative methods that use expert knowledge and quantitative methods that employ historical data. Most forecast problems require a mixed approach to ensure the best possible result. Thus, historical data, knowledge of future events that might impact the forecasts, and experimental knowledge are the basics of any good forecast [4, 5]. Any forecast is based on the idea that the considered factor will have a similar impact over time and that no major unpredictable event, such as natural disaster, will happen.

This master thesis was developed in collaboration with *NOS SGPS* in the scope of investment in analytic, development of useful and necessary tools in health. *NOS SGPS* has a highly important partnership with *Lusíadas Saúde SGPS, SA* that includes several solutions from mobile communications to user experience. In coordination with *NOS SGPS* digital innovation and transformation team, this project was developed with the *Lusíadas Saúde SGPS, SA* to expand the offered solutions. The *Lusíadas Saúde SGPS, SA* has high health and quality standards that cannot be compromised. The ED, by its nature, is always the department with most uncertainty and therefore, the hospital is investing in solutions that reduce it, improve health, and do not waste resources.

In this work, a forecasting tool, capable of forecast the ED visits, will be developed, using literature insights, health professional knowledge, and historical data. This process will be applied to *Hospital Lusíadas Lisboa* (HLL) and it will be tested in a real-life situation.

2. Methodology

This work aims the construction of a predictive tool capable of forecasting the influx of patients to an ED as accurately as possible. Although the use of predictive techniques in ED forecasting is well documented in the literature, there is a lack of methodologies that consider the knowledge of health professionals. The proposed methodology is established in 3 main pillars: literature review, workshops/interviews with health professionals responsible for the ED at the hospital, and historical data. This methodology is designed to be used in ED visits forecasting and complement the existing mathematical methods with the crucial knowledge of health professionals.

A flowchart of the proposed methodology and the respective main steps are given in Figure 1, where the yellow line corresponds to the steps using mainly the historical data, the green lines the importance of literature in that specific step, and the purple lines are the intervention of hospital staff in that step. Additionally, the black lines express the merge of the three main pillars.

Next, the implementation of the methodology to the HLL will be discussed. Although this methodology was applied to daily and hourly data aggregation, only the first will be presented.

2.1. Research and Context

In this step, the relevant factors encountered in the literature, such as weekly seasonality, should be sum up and their importance discussed with health professionals. Essentially, this part of the process requires the insights of management personnel, preferably responsible for managing the ED, about the main factors that can influence visits. The literature review was based on the research of keywords like "health forecast" or "ED visits forecast" in search engines with scientific manuscripts, such as "Web of science" or "IEEE Xplore Digital library". Meetings with Engr. Filipa Marques, responsible for the management of Imagiology, Emergency Departments, Cardiology, Nuclear Medicine, Oncology, Specialty Exams, and Pain Unit at the HLL, were crucial to gather the knowledge of healthcare professionals. The first meeting consisted of the presentation of the project, on the decision of the hospital in the group that would be fit, and discussion of the main necessities verified by this institution. In this context, it was determined the study of the Adults Care Unit (ACU). Engr. Filipa Marques indicated the weekly seasonality and the strong relation of ED demand to holidays, as factors to be studied. A workshop was developed, in cooperation with HLL, where, Doctor Sofia Lourenço, the ED medical responsible, Nurse Vanda Gomes the ED nursing responsible and Dulce Monho, the ED production manager responsible at HLL were present. This workshop main objectives are:

- Introduction to the project objective;
- Presentation of the work already developed;
- Discussion of factors that may have an impact on ED affluence;
- Discussion of data problems and inconsistencies;
- Validation of proposed solutions/actions;
- Discussion and validation of accuracy measures.

These professionals are responsible for the shift distribution in the department. In this context,

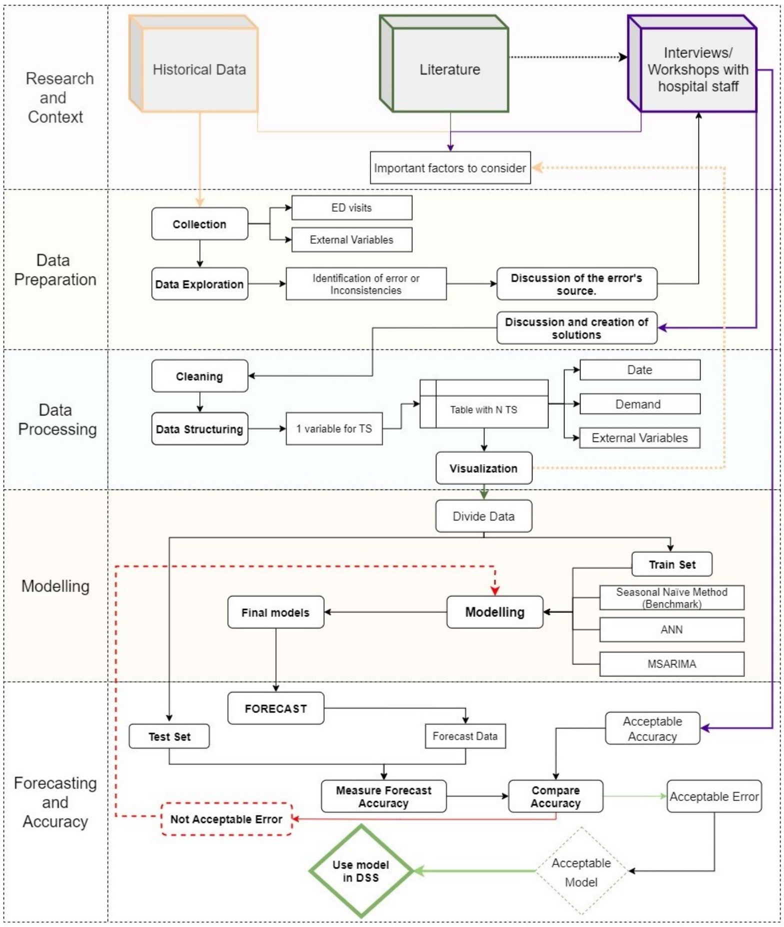

Figure 1: Proposed methodology flowchart.

their opinion regarding the best data aggregation in a forecast, the most effective accuracy measures, and what are tolerable errors, meaning errors that do not compromise the safety of the patient or the quality standards of the group, were taken into account. From that discussion we reach the following conclusions:

1. Holidays, such as Christmas have a great influence on ED visits;

2. Hourly forecasting is the main goal; however, it is preferable to have an accurate daily forecast than an hourly forecast;

3. MAE can be used as a reference accuracy measure since it is important to know the absolute error. The RMSE can be used as secondary criteria to penalizing models that have extreme errors;

4. The acceptable MAE was established as 10 patients for a daily forecast.

The inside knowledge is crucial; however, it does not replace the use of historical data. It was determined the use of data between May 2017 and December 2019 [6]. At this point, in the methodology, the health forecast principles should be remember [7, 8] and verify if all are defined: **Focus** - forecast the ED visits at the HLL; **Data aggregation** - daily forecast (test the hourly forecast); **Horizons** - 3 months; **Uncertainty and error** - the MAE inferior to 10 patients (in a daily forecast).

2.2. Data Preparation

In this phase, it is necessary to have well-defined objectives, the type of data needed, and have established study period. Subsequently, the data from the hospital database and the external data from other sources that may be relevant, for instance, a weather institute, should be gathered.

The HLL collects the data correspondent to every emergency episode in the ED for administrative and medical purposes. Thus, the data used in this work was not gathered with this purpose. The data collection was made by hospital personnel and a "xls" file, containing all emergency episodes registered at ED between $2017-05-01$ and $2019-12-31$, was made available by the IT department [6]. The first examinations were conducted using *Microsoft Office Excel* and only Date, Hospital Arriving Time (HAT) and Triage Time (TT) were used to make the process more efficient. The meetings with HLL professionals confirmed the factors suggested in the literature and added two more important factors, for instance, the weather and holidays. Therefore, data regarding influenza in Portugal, all holidays in Portugal and Lisbon, and weather conditions in Lisbon for the considered period have to be collected.

Holidays: In Portugal, there are 14 holidays during the year that need to be considered. We created 2 variables to deal with the possible holiday effect. The first one is called "Days After Holiday", to know how many days have past since the last holiday. Thus, in this TS, the value 0 corresponds to the day of a holiday, the next day has value 1 and so on. All holidays are considered in this TS. The second is called "Difference" and its values correspond to the difference between the holiday in question and the average demand for that normal weekday, represented by equation 1. This is applied to the 11 fixed holidays, their previous and following days as well.

$$\text{Difference of EE}_t = \text{EE}_t - \bar{EE}_{wd,y} \quad (1)$$

Weather: It is important to have hourly weather data for LMA [9]. The data has four columns: datestamp, the temperature in Celsius, relative humidity (%), and the precipitation in millilitres. To deal with the missing values of temperature, relative humidity, and precipitation, it can be used the average of the previous three hours for each variable

Influenza: Lastly, the influenza activity in Portugal during the study period [10]. The available data is the national daily ratio between the number of emergency episodes with influenza, flu syndrome, or other respiratory infections and the total number of emergency episodes. The required data is not available for most of 2018 and only data from January and December are available. Therefore, an average between the values of 2017 and 2019 was used for the missing values in 2018.

Then, the data exploration exposed some problems, such as coverage, completeness, or accuracy, that can not be ignored during the analysis. The hospital data, that was collected, has three main problems, for instance, missing or abnormal values especially in the time columns, impossible data combinations, and missing information. These mistakes can have diverse causes, such as bad transcriptions, errors during filling the admission forms of patients, or problems with software synchronization since HLL uses two software in the ED. The episodes with errors, where it was impossible to determine the HAT, were excluded as well as the ones with HAT and TT incompatible.

2.3. Data Processing

This phase is essential since some errors can prevent the creation of a model. For example, most of them cannot deal with missing values. The recommended data structure consists of a column for the date (or timestamp in case of sub-daily data), one column for demand, and one column for each external variable. This guarantees that each column is an independent TS, which starts and ends at the same time of the same day. The time interval between observations must be constant.

The cleaning processes excluded 0.23% of the original data. The remaining data was converted into the recommended data structure. The data analysis was not only consistent with the information obtained from hospital professionals but also the HLL data presents a similar weekly and monthly seasonality as the ones described in the literature. Applying correlation equation [11] to this data, it was demonstrated that the demand has a strong correlation with weekday, influenza-tax (influenza ratio), and a less important relation with temperature and difference. These values are relevant to decide which external variables will be used in the modelling process.

2.4. Modelling in R

The construction of a predictive model using historical data happens during the **Modelling** phase. Supervise learning is used to identify the best models since it is possible to determine the out of sample accuracy. As illustrated in Figure 1, to apply this technique, it is necessary to divide the data into training and testing sets. The models are trained using the first set of data and then, forecast for horizon correspondent to the testing set. Using the forecast values and the real ones, the accuracy measures can be calculated and the models' accuracy assessed. In this project we used R, that is a "free software environment for statistical computing and graphics", particularly powerful for forecast [12]. To achieve the proposed objective, some prediction tools, such as the ones present in fpp3 [13], fabletools [14] and fasster [15] packages, were used.

Since the data has strong seasonality patterns and also presents trend, it can be only considered methods that take at least one of those characteristics into account during the analysis. Thus, the Seasonal Naïve Method (SNAIVE), Exponential Smoothing State Space Model (ETS), Autorregressive Integrated Moving Average model (ARIMA), Neural Network Time Series Forecasts (NNETAR), and Fast Additive Switching Of Seasonality, Trend And Exogenous Regressors (FASSTER) should be tested [13–15]. Only ARIMA, NNETAR, and FASSTER can handle exogenous regressors, such as temperature thus, several combinations of these were tested for each function. The model specifications are always expressed in the same form: *response terms*. The left side corresponds to the response variables and on the right side, the formula of the model is specified. This structure includes *specials*, for instance, *trend()* or *season()* that can be used. They differ from model to model. The methods are applied to the training set automatically, meaning that the program learns and determine the best model parameters for each class. The training set starts on $2017-05-01$ and ends on $2019-09-30$, while the testing sets start on $2019-10-01$ and ends on $2019-12-31$. The data used to test the model corresponds to the last quarter of 2019, where there are five national holidays. This means that about 15 of 92 days are an exception to normal weekdays since they are before, after or a holiday.

2.5. Forecasting and Accuracy

In the modelling phase, various methods with different features are used to create models that explain different seasonality patterns. They can be seen as time-dependent functions that better describe the data. If a model uses one or more external variables, the same principle can be applied, but with extra variables. Since several methods applied to model data, the same requires criteria to compare their performance. There are several accuracy measures, such as Mean Absolute Error (MAE) [16], Mean Absolute Percentage Error (MAPE) [17], and the Root Mean Square Error (RMSE) [16] that scale the magnitude of the error. While MAE and RMSE are scale-dependent, the MAPE is not.

The models, produced in the previous phase, are used in the *forecast* function with the external data file to produce the intended forecast. Then, the accuracy measures can be employed to compare models. Another way to make sure that the model is the best possible is to make residual diagnostics. To used it, the residuals should be calculated. If their mean is 0 and there is no correlation between them [18], the model has retrieved all the information present in the data and the residuals are white noise.

2.6. Decision Support Tool - ForcastER

A solver-oriented DSS was implemented in this project, aiming the efficient use of forecasting tools by the medical or management team. It was developed using the same programming language. Besides the packages mentioned above, the shiny dedicated packages, such as shinny and shinydashboard [19] were used. This is a simple, visual, and interactive interface between the R code and the user. It guarantees that any person can use the tool without any knowledge of R. A Shiny App has two main parts: the *UI* and *server*. The user interface (UI) is responsible for the layout, creating all input and output variables, while the server part is responsible for all interactions and transformations between the input and outputs. The following questions should be in mind to assure the best possible design and operation of the DSS:

- **Who needs it and why?** - ED management team at HLL to allocate human resources to the ED (minimize costs);
- **What are the advantages?** - Easy implementation of forecasting tools crucial for the work distribution;
- **Where does it fit in the business process?** - Planning of resources (especially human) for the ED.

The Shiny App, developed for this project, has 3 main sections, for instance, "Home" (Figure 2), "Daily Forecast", and "Hourly Forecast". The first page has the instructions for the use of both types of forecasts. In the "Daily Forecast" and "Hourly Forecast", the user can introduce the necessary inputs to produce a forecast. Each page is responsible for a type of data aggregation thus, the necessary inputs are slightly different and are summarized in the user guide (home page).

The necessary data, in each CSV file, varies with the external variables and with the type of data aggregation the user chooses to use. Due to the DSS implementation, the columns must have the exact same names presented on the home page. The DSS has a set of instructions to produce the forecast but it has no data. For safety and efficiency reasons, the user has to upload the data to the Shiny App. The data is lost when the App is closed. All the files upload to the DSS should comply with the instructions above, otherwise, the DSS will not produce meaningful results or not work at all. It should be used to forecast the demand of an ED with similar behaviour to the ACU. This will assure that the models used in this DSS are adequate and it will produce reliable forecasts. This DSS is planned to assist the decision-making process without suggesting any course of action. It will produce, display, and export the forecast values of the requested period using the chosen variables.

3. Results

The models used are discussed in section 2.4. Here, the results of the best models for each data aggregation will be addressed. As mentioned previously, the original data was arranged by daily demand, where each line corresponds to a day. The data set was imported to R.

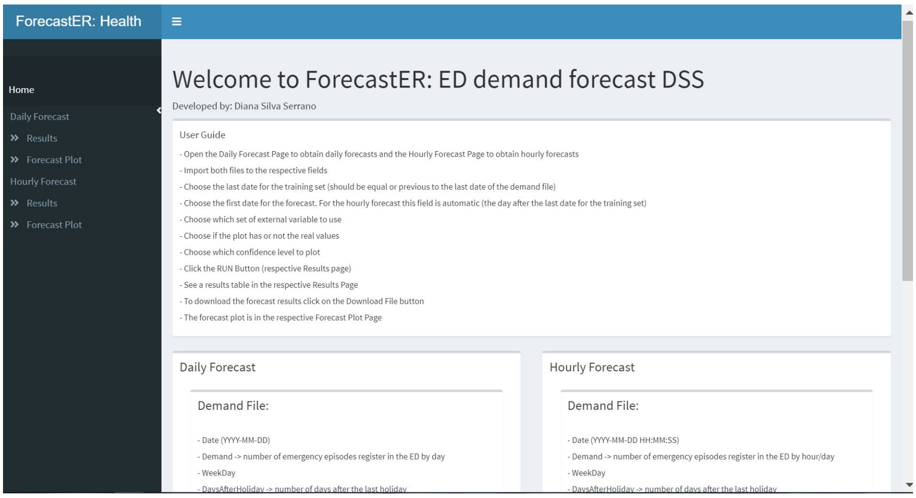

Figure 2: Home page of the DSS developed for this project.

This analysis is conducted in a controlled environment where all variables are accessible. However, in a real situation, it is not possible to have future values for some variables, such as temperature, perception, or humidity. It can be challenging to have forecast models that depend on future values of variables that are not known or easy to find. On the other hand, it can be easily determined the future values of variables, such as weekday or holidays. Consequently, if it is possible, it should be prioritized models that are equally good and do not require this type of variables. The professionals were consulted to define the criteria needed to choose the methods:

- MAE < 10 for testing set;
- Minimize RMSE and MAE for testing set;
- Minimize computational time;
- Minimize the number of external variables needed.

Table 1: Name, MAE, MAPE (%), and RMSE of the best models trained for the daily data aggregation. The accuracy measures for the out of sample forecast from October to December 2019.

Name	MAE	MAPE	RMSE
fasst_influenza_tax	9.47	7.23	12.20
fasst_influenza	9.32	7.05	12.30
fasst_influenza_10	9.61	7.29	12.40
fasst_work_holi	9.54	7.39	12.40
fasst_work_flu_10	9.61	7.32	12.40

First, it should be identified methods to minimize RMSE, MAPE, and computational time. The quartiles of the absolute errors of each model as well as the interquartile range (IQR) were calculated and used as a tool of differentiation. Therefore, all models with a computational time inferior to 152.23 seconds, MAE < 10, MAPE < 7.73, and RMSE < 12.98 in the testing set, should be considered to be used in the DSS. The models that did not comply with those values or use any weather variable were excluded. The remaining are present on Table 1. In Figure 3, it is possible to see the graphical representation of the forecasts produced using these models and conclude that, although they have similar accuracy measures they have different behaviours. All of the real values are inside the 80% confidence interval of all models. Thus, they will be applied to the final DSS.

4. ForcastER

It is important to test the accuracy of these models in a real forecast. Therefore, we used the DSS to forecast for the first quarter of 2020. It applies the models chosen in the previous section. Time, Weekday, Days After Holiday, and Difference are mandatory in any kind of forecast. Therefore, a file with dates and the value of these variables is always required in this DSS. The user selects the set of external variables to be analysed and the DSS runs all models that contemplate at least one of those variables. The default option will run all models (if the variables are present in the files). The demand data, corresponding to the first quarter of 2020, was retrieved from the HLL database to test the DSS performance [6].

For a daily forecast of the first quarter of 2020, the DSS needs the historical data file and the external data, such as influenza or workday for the dates to forecast. The last day of the training set $(2019-12-31)$ should also be chosen as well as the first day of the forecast $(2020-01-01)$ in the appropriate fields. Next, the set of external variables to use in the forecast should be chosen. Lastly, it is needed to decide if the plot should represent the confidence interval of each model or not and if it should plot the real values or not. The program

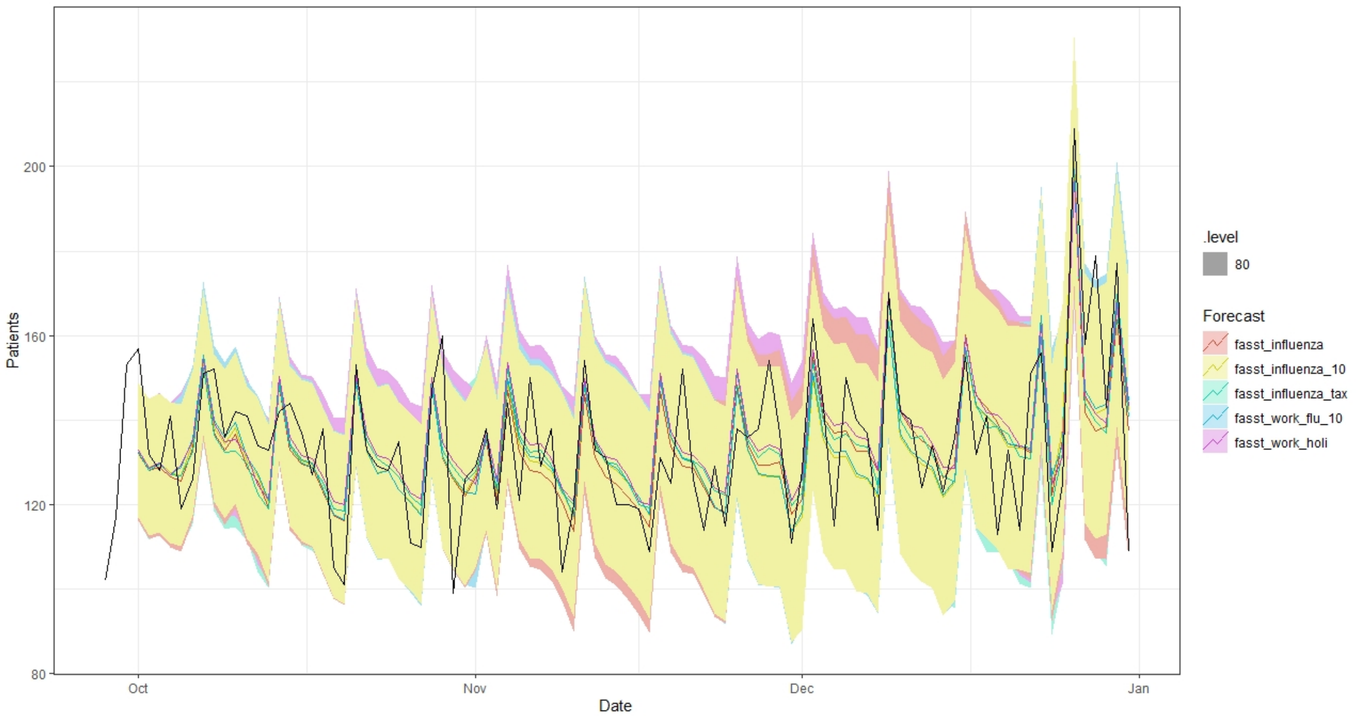

Figure 3: Real number of daily attendance to the ED in black [6]. Forecast values and respective confidence interval of the 5 best models. These values correspond to forecasts of the ACU demand between October and December 2019 at HLL.

will start the forecast when the button "RUN" is clicked (results page).

As can be verified in Figures 5, the model is working well until the middle of march, where the real values are not near the forecast ones. The daily and hourly forecast have the same discrepancy in March; however, only the daily data aggregation is used in the analysis to discover the origin of this discrepancy.

To confirm the origin of this discrepancy, the same period (March) was forecast, using the entire set (until February twenty-nine 2020). An exhaustive examination of the ED demand in the first quarter of 2018 and 2019 was also made. With these examinations, it is pretended to determine if the variance is due to the forecast horizon, the month in question, or if the model is not working properly. To evaluate if the March demand is typical or has meaningful differences from previous years, the relative difference between the normalize demand of two consecutive years (for each day) was calculated. The $D_{norm_{d,m,y}}$ represents the daily demand without the weekday factor. Thus, it can be compared calendar days of consecutive years without the impact of the weekday on the demand. To normalize the daily demand relative to the weekday, it was determined the average weekday demand for each month of each year. In Figure 4, it can be seen the graphical representations of the values obtained. The differences between 2019 and 2018 belong the range $[-25\%, 23\%]$, while, for the same period, the differences between 2020 and 2019 belong to the range $[-136\%, 56\%]$. As it can be observed in Figure 4, the changes are more radical during March. The forecast of March, using data until February twenty-nine 2020, improves the accuracy measures for all models. Using both of these investigations, it can be inferred that the errors exhibited by the DSS during the forecast of March, are not associated with the models.

The most notorious variances started in March. This can be easily explained by the outbreak of the SARS-CoV-2 that is responsible for the COVID-19 [20]. The classical definition of a pandemic is "an epidemic occurring worldwide, or over a very wide area, crossing international boundaries and usually affecting a large number of people" [21]; however, the WHO has also evaluated the population immunity, virology and the severity of the disease before declaring it as a pandemic. COVID-19 is a respiratory disease that affects everyone, being particularly severe for the elderly. There is no vaccine or treatment. The WHO advice to delay the COVID-19 spread is by self-isolate, social distance, and washing hands several times a day [20].

To better understand and evaluate the impact of this pandemic in the ED demand, the timeline of these events should be studied. The WHO declared that "COVID-19 can be characterized as a pandemic" on March 11, 2020, [20]. A few days later, on March 18 2020, the *Presidente da República Portuguesa* declared, trough the *Decreto do Presidente* *da República n.º 14-A/2020, Estado de Emergência* [22]. The government use it to enforce the quarantine advise by national and international health authorities. These extreme measures aimed the delay of the epidemiological curve and prevent the overflow of patients in an ED or hospital. In Portugal, only essential services were operational, such as supermarkets, banks, or pharmacies, and all citizens were asked to only leave home if it was extremely necessary, particularly elderly citizens. Infected patients and people under vigilance by the DGS were not allowed to leave home. It was the first time, since 1974, that Portugal was in *Estado de Emergência* and the first time in Democracy that Portuguese have their movements limited. Comparing the dates highlighted in this timeline and the demand differences in Figure 4 it can be inferred that it is a possible causality relation. This indicates that the COVID-19 outbreak causes the abnormal demand pattern present in March 2020. This is supported by the fact that the DSS has better accuracy one the training set involves January and February 2020. This happens because this months already present a different tendency and the model recognizes the change in the ED demand behaviour.

5. Discussion

The methodology was tested in the ACU at HLL and the accuracy measures were better than the ones found in the literature. Comparing them, it is important to highlight that the results presented for each work are the best results for each method in that study. For the daily forecast, the MAPE obtained by an ARIMA model in *Sun, Yan et al.* [23] has better accuracy than the models in this work; however, the latter are better than the ones presented in *Jones, Spencer S. et al.* [17], *Marcilio, Izabel et al.* [24], and *Calegari, Rafael et al.* [25]. Even when forecasting the daily demand for 3 months, these models are still better (excluding the model of *Sun et al.*). In the DSS, the models present MAPE results similar to the literature. Therefore, it can be concluded that the models produced using the proposed methodology have better performances when compared to the literature.

According to the opinions of the professionals, there were external factors, such as holidays and weather variables, that influenced the ED demand. The weekly seasonality and the influence of the flu season were also designated as important aspects to analyse. In the literature, these are presented as important and sometimes decisive to obtain a reliable forecast model. It is commonly accepted the impact of calendar variables, such as weekday or holidays, on ED demand [23, 24]. However, in the literature, there is no consensus regarding the importance of weather variables. *Jones, Spencer S. et al.* [17], *Marcilio, Izabel et al.* [24], and *Sun, Yan et al.* [23] did not find any significant relation between weather variables and variations in ED demand. These variables also have limited benefit in ED demand forecasting tool. On the other hand, *Holleman, Donald R et al.* [26], *Jones, Simon A. et al.* [27], and *Shah, Sparsh et al.* [28] found this relation important. These studies demonstrated an important relationship between the weather and the ED affluence. For example, the inclusion of variables related to snowfall improves the accuracy of the model [28]. The divergent conclusions in these studies can be associated with the type of weather for each city, the social behaviour of the populations, or the nature of its characteristic diseases. This work did not examine any other alternative type of relation between the ED demand and other variables, such as the square temperature or a non-linear relation. However, there is a correlation between weather variables, such as temperature or humidity, and the accuracy of the models found. In a real situation, the DSS user will not have the temperature, precipitation, or humidity for the forecast horizon. Therefore, as mentioned previously, the nature of the weather variables implies a limitation to their use in predictive tools. Nevertheless, several reliable short-term weather forecasts could be used in short-term ED demand forecast. These would allow the use of the best models and variables without compromise the viability of the forecast.

6. Conclusions

The increase in ED demand in recent years and the health hazards associated with long waiting times, impose the necessity of forecasting the ED demand. These are particularly severe in private hospitals since the demand has increased by 68.5% between 2012 and 2018, in the LMA [1]. Therefore, the main objective of this work is focused on the development of an ED demand predicting tool. The developed methodology contemplates two types of knowledge. It aims the use of experts knowledge to maximize the outcome of mathematical frameworks. Its first application produces better results than the ones available in the literature, as discussed previously. Although most studies prove a relation between weather variables and ED demand, some authors argue that the benefit of incorporating them is not significant. During the extensive analysis in this work, it was demonstrated the influence of external variables, such as temperature, in the accurate prediction of the ED demand. However, the possibility of using these factors in real-life situations is limited. The project accomplished the objectives agreed with the hospital and produced a tool capable of predicting the ED demand. This tool is based on forecasting methods that, by definition, do not

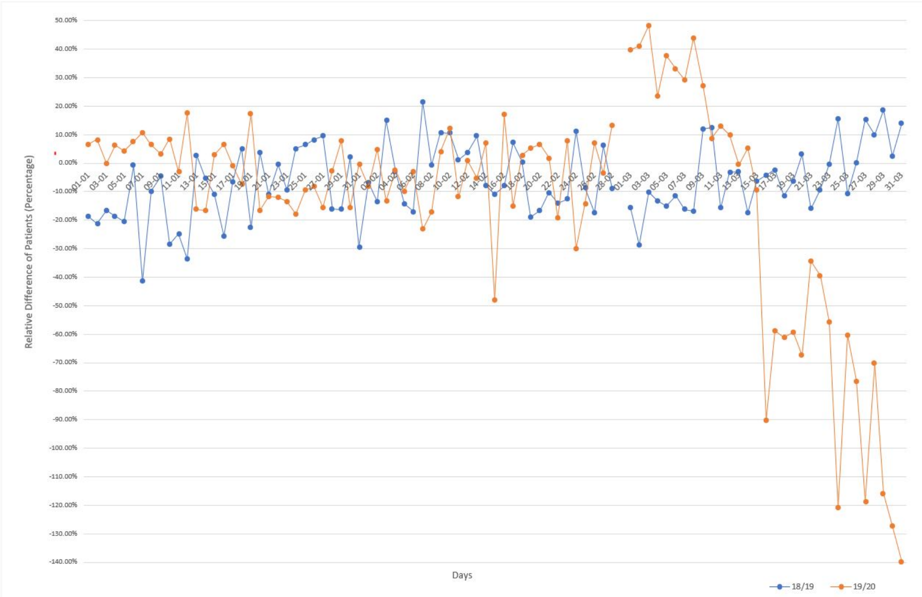

Figure 4: Relative difference of emergency episodes register between two consecutive years for the same calendar day at the ACU [6].

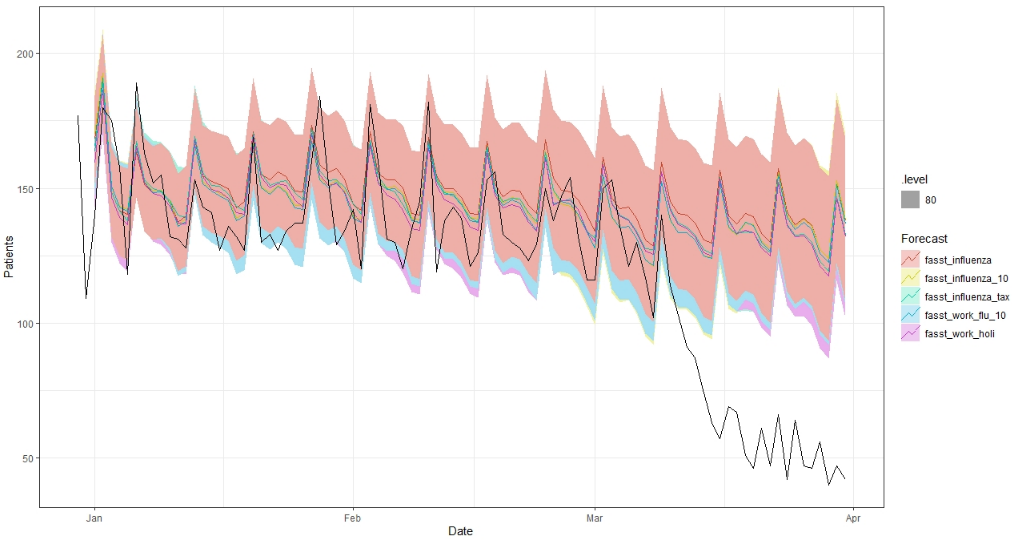

Figure 5: Real number of daily attendance to the ED in black [6]. Forecast values and respective confidence interval of the 5 best models. This values correspond to the demand forecast for the *Hospital Lusíadas Lisboa* ACU between January and March 2020.

consider unpredictable events, such as pandemics. Thus, is natural that the DSS cannot predict the impact of such an event on ED demand.

Forecasting methodologies can be extremely helpful in health, especially in ED management. The health demand has been increasing, emergency episodes increase every year and hospital need to be able to allocate the necessary resources. Patients have higher standards every day, particularly in private hospitals. They see waiting times in an ED as health hazards and will do everything to avoid them. Thus, the capability to forecast and adjust resources will be increasingly crucial to manage an ED. The constant development of methodologies and predicting tools, in health, is vital to continue to meet the increasingly high-quality standards of patients.

Future Work The work developed has some limitations, such as the consideration of only one division of an ED in one hospital. The methodology should be tested using other ED, preferably with different characteristics, such as size, served population, or public hospitals. Similar accuracy values would confirm the improved performance over literature methodologies. Further, this tool only uses the methods that produce the best models, tested in the modelling phase. Thus, the possibility of update the methods accordingly to the data should be further studied. This would allow the use of the tool in the remaining ED divisions of the HLL or even ED of other hospitals. Future work should also investigate the advantages of this tool (or a similar one), being incorporated in the hospital management system. This would enable the use of real-time data that could be translated into the real-time forecast. In this case, the tool could provide forecasts for the next few days and/or for the following weeks that contemplate the real demand. Using those capabilities, a DSS can be incorporated in the software to help organize and decide the best schedules of doctors or nurses. It could also be essential to manage not human resources essential in an ED. This implicates a great amount of research; however, it can be a decisive tool to help manage human resources efficiently.

Acknowledgement

The work presented in this thesis is a partnership between Técnico, NOS SGPS (Lisbon, Portugal), and Lusíadas Saúde SGPS, SA during the period September 2019 to June 2020, under the supervision Prof. Mónica Oliveira at Instituto Superior Técnico. The thesis was co-supervised at NOS SGPS by Miguel d'Aguiar.

References

[1] INE, "Estatísticas da Saúde." [Online]. Available: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=257793024&PUBLICACOESmodo=2

[2] "Manchester Triage System — definition of Manchester Triage System by Medical dictionary," 2019. [Online]. Available: https://medical-dictionary.thefreedictionary.com/Manchester+Triage+System

[3] Serviço Nacional de Saúde, "Atendimentos em Urgência Hospitalar por Triagem de Manchester," 2020. [Online]. Available: https://transparencia.sns.gov.pt/explore/?sort=title&q=Triagem

[4] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Melbourne, Australia: OTexts, 2019. [Online]. Available: https://otexts.com/fpp2/

[5] P. E. Gaynor and R. C. Kirkpatrick, *Introduction to Time Series Modeling and Forecasting in Business and Economics*. McGraw-Hill College, 1994.

[6] "Hospital Lusíadas Lisboa: base de dados," Lisbon, 2020.

[7] I. N. Soyiri and D. D. Reidpath, "An overview of health forecasting," pp. 1–9, 2013.

[8] G. Zotteri, M. Kalchschmidt, and F. Caniato, "The impact of aggregation level on forecasting performance," in *International Journal of Production Economics*, Attila Chikan, Ed., 2005, pp. 479–491. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092552730400266X

[9] "Weather archive in Lisbon." [Online]. Available: https://rp5.ru/Weather_archive_in_Lisbon

[10] Ministério da Saúde, "Atividade do Síndrome Gripal nos Cuidados de Saúde Hospitalares," 2017. [Online]. Available: https://transparencia.sns.gov.pt/explore/dataset/atividade-sindrome-gripal-csh/table/?disjunctive.ars&disjunctive.indicador&sort=periodo

[11] "CORREL function - Office Support." [Online]. Available: https://support.office.com/en-us/article/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92

[12] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2019. [Online]. Available: https://www.R-project.org

[13] R. Hyndman, RStudio, M. O'Hara-Wild, and G. Athanasopoulos, "fpp3 package — R Documentation." [Online]. Available: https://www.rdocumentation.org/packages/fpp3/versions/0.2

[14] M. O'Hara-Wild, Hyndman Rob, and E. Wang, "fabletools package." [Online]. Available: https://fabletools.tidyverts.org/index.html

[15] M. O'Hara-Wild and R. Hyndman, "FASSTER Package." [Online]. Available: https://github.com/tidyverts/fasster

[16] K. S. Dadu and P. C. Deka, "Multistep Lead Time Forecasting of Hydrologic Time Series using daubechies wavelet - neural network hybrid model," *International Journal of Scientific & Engineering Research*, vol. 4, no. 10, 2013. [Online]. Available: http://www.ijser.org

[17] S. S. Jones, A. Thomas, R. S. Evans, S. J. Welch, P. J. Haug, and G. L. Snow, "Forecasting daily patient volumes in the emergency department," *Academic Emergency Medicine*, vol. 15, no. 2, pp. 159–170, 2 2008.

[18] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia.: OTexts, 2019. [Online]. Available: https://otexts.com/fpp3/

[19] "Shiny - Function reference," 2020. [Online]. Available: https://shiny.rstudio.com/reference/shiny/1.4.0/

[20] "WHO — World Health Organization," 2020. [Online]. Available: https://www.who.int/

[21] J. M. Last, *A Dictionary of Epidemiology*, 4th ed., R. A. Spasoff, S. S. Harris, and M. C. Thuriaux, Eds. New York: Oxford University Press, Inc., 2001.

[22] "Decreto do Presidente da República n.º 14-A/2020 - DRE," 2020. [Online]. Available: https://dre.pt/web/guest/home/-/dre/130399862/details/maximized

[23] Y. Sun, B. H. Heng, Y. T. Seow, and E. Seow, "Forecasting daily attendances at an emergency department to aid resource planning," *BMC Emergency Medicine*, vol. 9, 1 2009.

[24] I. Marcilio, S. Hajat, and N. Gouveia, "Forecasting daily emergency department visits using calendar variables and ambient temperature readings," *Academic Emergency Medicine*, vol. 20, no. 8, pp. 769–777, 8 2013.

[25] R. Calegari, F. S. Fogliatto, F. R. Lucini, J. Neyeloff, R. S. Kuchenbecker, and B. D. Schaan, "Forecasting Daily Volume and Acuity of Patients in the Emergency Department," *Computational and Mathematical Methods in Medicine*, vol. 2016, 2016.

[26] D. R. Holleman and R. L. Bowling, "Predicting Daily Visits to a Walk-in Clinic and Emergency Department Using Calendar and Weather Data," *Journal of General Internal Medicine*, vol. 11, pp. 237–239, 1996.

[27] S. A. Jones, M. P. Joy, and J. Pearson, "Forecasting demand of emergency care," in *Health Care Management Science*, vol. 5, no. 4, 2002, pp. 297–305.

[28] S. Shah, J. Murray, M. Mamdani, and S. Vaillancourt, "Characterizing the impact of snowfall on patient attendance at an urban emergency department in Toronto, Canada," *American Journal of Emergency Medicine*, vol. 37, no. 8, pp. 1544–1546, 2019. [Online]. Available: https://doi.org/10.1016/j.ajem.2019.05.032