

The 3rd International Workshop on Healthcare Open Data, Intelligence and Interoperability
(HODII)
October 26-28, 2022, Leuven, Belgium

Predictive analytics for hospital inpatient flow determination

Diogo Peixoto^a, Agostinho Barbosa^b, Hugo Peixoto^a, João Lopes^a, Tiago Guimarães^a,
Manuel Santos^{a*}

^aALGORITMI/LASI Research Center, University of Minho, Portugal

^bCentro Hospitalar do Tâmega e Sousa, Portugal

Abstract

Currently, the efficient planning of resources in hospitals present a responsibility of extreme importance in the management of the various clinical units. In Intensive Care Unit (ICU), a hospital service where patients require constant observation and control, considering the high costs incurred with hospitalized patients, the optimization of these factors assumes an extremely important role. Given its unpredictability, this study focused on a characterization of this unit, identifying existing patterns, during a 5-year period, 2017 to 2021, at the Centro Hospitalar do Tâmega e Sousa (CHTS), providing a set of useful information crucial for decision making. Additionally, a prediction of future ICU admissions is performed using time series and Machine Learning (ML) models. However, the models did not reveal a predictive ability with an adequate level of reliability.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the Conference Program Chairs

Keywords: Inpatient Flow; Machine Learning; Predictive Analytics

1. Introduction

Currently, hospitals store a large volume of data, which leads to the need to apply techniques capable of extracting useful information for the decision-making process. These techniques have been presenting applications with high impact on the improvement of care provided to patients, given the various existing applications in the health sector [1].

* Corresponding author.

E-mail address: mfs@dsi.uminho.pt

Emphasizing Intensive Care Units (ICU), one of its inherent problems is associated with the fact that hospitalizations are not planned, leading to a constant lack of knowledge. Given this situation, this study is aimed at the knowledge of the most common admissions to the Polyvalent Intensive Care Unit (UCIP) of the Centro Hospitalar Tâmega e Sousa (CHTS), and a detailed characterization of the same, to allow a more adequate selection of admitted patients, since this is the main cause of a deficient occupation of hospital resources. Likewise, Data Mining (DM) techniques were applied to predict the future admission of patients to the UCIP, so as to allow a more optimized management of this clinical unit.

2. Background

2.1. Resources Planning in Hospital Settings

Currently, the continuous overcrowding associated with hospitals is notorious, which is caused in most situations by a lack of beds. Consequently, besides the degradation of the quality of service provided, it can lead to the postponement of pathological treatments and to an increased risk of contracting contagious diseases [2]. Very recently in Portugal, because there was no vacancy in the Neonatology Service of Santa Maria Hospital, a pregnant woman died while being transferred to another hospital due to cardiorespiratory arrest [3]. A situation that could have a different outcome if there was a more assertive management, prepared for the existence of unpredictability. However, the opposite also occurs, resulting in unnecessary costs in terms of resources, since the needs are smaller than the number of beds available. Hospital bed management is a present problem in hospitals, which, when executed incorrectly, causes a mismatch between means and human resources, implying considerable costs.

Taking all these aspects into consideration, if it is possible to estimate with a considerable level of accuracy the number of patient admissions using DM techniques [4], it becomes feasible to assist professionals in developing more assertive planning.

2.2. Related Works

Currently, ML techniques applied to Intensive Care Units have gained a growing interest to improve their management [5]. Given the presence of unplanned admissions inherent to this unit, the forecasting of future admissions is of particular interest, as it enables a more optimized management of resources, both financial, material and human.

In 2022, Alshanbari et al. [6] focused their study on classifying patients with COVID-19 who aim to require ICU admission, based on demographic and clinical data collected between May 2020 and January 2021. Among several classification techniques, the Support Vector Machine (SVM) was the one that presented the best performance, with an accuracy of 88.1%. Thus, this study may provide a more optimized bed management, since it allows for a real-time determination of priorities [E1].

Taking into account other services, Adriana Vieira [7] developed a Master's dissertation in Statistics, at the University of Minho, which consisted in predicting the number of patients admitted daily in three emergency services in the Hospital of Braga, Portugal and the number of patients who, when admitted to the emergency service, will be hospitalized. A dataset was used regarding daily patient admissions to the emergency department from the beginning of 2012 through the first quarter of 2016. Other variables taken into consideration were the day of the week, month of the year, holidays, the minimum and maximum temperature of the previous day, the number of flu visits made during a week, and the concentration of some pollutants. Despite being statistically significant, the non-seasonal variables showed little predictive value. Thus, regarding emergency room admissions, multiple linear regression techniques were used and a temporal correlation was found (7 days for the general emergency room and 8 days for the others) [E2].

Also concerning Braga's hospital, Miguel Silva [8] compared some models in order to select the best model to predict the number of users admitted daily in the emergency department of Braga's hospital and to test the relationship of environmental variables with the patients' arrivals. In this research, it was also highlighted the fact that holidays, minimum, average and maximum temperatures on the day of the user's arrival and precipitation

obtained a weak correlation. Regarding the forecast models, the one with the best results was the seasonal ARIMA with an average absolute percentage error of less than 6%. [E3].

Considering the State-of-the-Art presented, the importance of predicting the number of future admissions in this unit should be emphasized, as it is an essential indicator in its management, since the use of these models provides a better planning of hospital resources. In general, we can conclude that exogenous variables do not have great predictive value for the problem in question, however, to prove these conclusions, this project also used this type of variables. In this way, all the information gathered in the state of the art will be crossed, to analyse which are the implications in the admission to the UCIP, following a regression approach, as of time series.

3. Materials and Methods

For this research, the DSR (Design Science Research) was applied, popular in Information Systems, allowing new technical and scientific knowledge to be acquired from the design of innovative artifacts, to solve a practical problem in a specific context. It is divided into six activities [9], which are: Understanding the Problem (1), Suggestion (2), Development (3), Evaluation (4), Conclusion (5) and Communication (6).

The CRISP-DM (Cross Industry Standard Process for Data Mining) methodology was also adopted, since it support the life cycle of a DM project. This is composed of 6 stages [10]: Business Understanding (1), Data Understanding (2), Data Preparation (3), Modeling (4), Evaluation (5), and Implementation (6).

4. Case Study

This section presents the processes and decisions taken in the case study. As previously mentioned, this project aims to predict future admissions to the UCIP of the CHTS inpatient service.

4.1. Business Understanding

To achieve improved efficiency in hospital bed planning and management by forecasting future admissions to the UCIP, this project was developed in collaboration with the CHTS. The selection of this unit is justified due to its unpredictability, as well as the fact that it is one of the specialties with the highest costs, both in terms of resources and money.

4.2. Data Understanding

The data collected are limited to a time interval of 5 years, from 2017 to 2021, alluding to admissions in the specialty under study and patient demographic data. Another data source, provided by IPMA, presents data regarding the maximum and minimum temperatures recorded daily. Using web scraping, the extraction of exogenous variables was performed. The analyses presented below come from the use of these same data.

Firstly, based on the patient's gender and Length of Stay (LOS), it was possible to verify that there is a greater presence of the male gender, however, this variable does not seem to have any influence on the LOS of the patients, since it is evident the similarity of the distribution of patients by the hospitalization time intervals in this specialty (Figure 1).

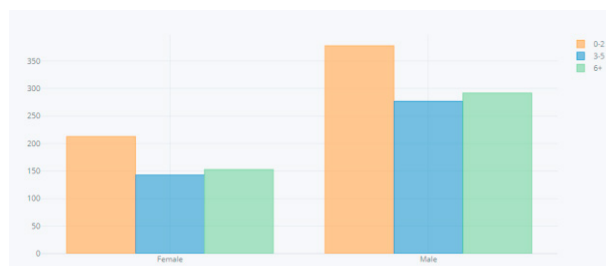


Figure 1. Number of patients by gender depending on LOS

Subsequently, highlighting the age range and keeping the focus on the LOS, it was found that most patients hospitalized in this specialty are between 40 and 80 years of age. Through Figure 2, it is possible to observe that, apart from the age group between 40 and 64 years, the time interval of less than two days of hospitalization is the one that occurs most frequently, and in the cases of patients under 16 years of age, none stays more than 24 hours in the UCIP.

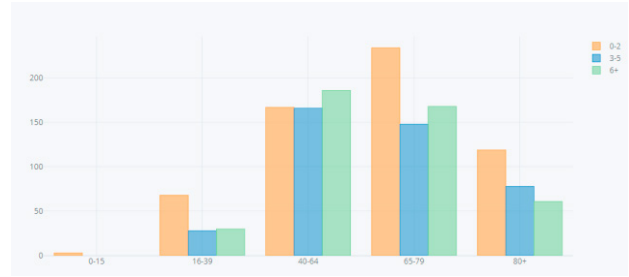


Figure 2. Number of patients by age depending on LOS

To understand the most frequent origins of patients admitted to this specialty, a graph-based study was developed, which allowed the identification of the previous specialty to which a patient had been admitted before being admitted to the UCIP, also identifying how the patient was admitted to the CHTS. Based on Figure 3, it was possible to conclude that the Emergency Department and the Surgery Specialty give rise to the most admissions in this unit.



Figure 3. UCIP's Inputs

Through Figure 4, the five most recurrent origins were identified, generating a graph with the distribution of hospitalization time intervals. It should be highlighted that, when the previous specialty is Vascular Surgery, the cases in which patients remain hospitalized for more than two days are rare. When the origin is Surgery, there is already a greater balance of cases, keeping the interval of 0-2 days more predominant, unlike the Urgency and UCIPSU where the LOS is higher.



Figure 4. Number of patients by inputs depending on LOS

Figure 5 shows that the Emergency Department continues to be the most recurrent, with high hospitalization times, despite the existence of a balance at this level, contrary to what happens in the Outpatient Department, which shows reduced hospitalization times.

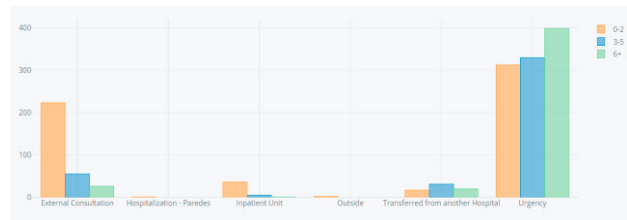


Figure 5. Number of patients by manner in which patients are admitted depending on LOS

4.3. Data Preparation

At this stage, to include most information as possible in the models, we proceeded to build a source of data relating to a calendar, contemplating the holidays, the designation of days of a week, and all festivities in the council. In addition, IPMA provided a set of records concerning the weather conditions in the time period under research, 2017 to 2021, where the variables were selected as maximum and minimum temperature, precipitation and

wind. All data sources were integrated to create a reliable data source with all the necessary information to serve as input for the models developed. These were coded so that they could be processed by ML techniques present in the Scikit-learn library.

4.4. Modelation

Based on the objective of this study, four regression techniques were applied, namely, Decision Tree (DT), Random Forest (RF), Linear Regression (LR) and Gradient Boosting (GB) and, also, a time series forecasting technique, Prophet (PH), as a term of comparison. The Cross Validation K-fold (CV) technique was used, since it allows the use of all data for training and testing, providing greater reliability in the models developed [11]. This study considered only one scenario, corresponding to the prediction of future daily admissions of patients to the UCIP, resulting in 5 models (1 scenario x 5 techniques).

4.5. Evaluation

To compare and evaluate the performance of the various models developed and verify if they meet the objectives previously determined, the evaluation metrics were identified, these being: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE) and R Squared (R2). Success criteria were defined for each defined evaluation metric:

- MAE, MSE, RMSE < 0,6;
- K2 $\geq$ 0,8.

5. Results And Discussion

Table 1 present all results achieved in the ML models.

Table 1. ML models results

	MAE	MSE	RMSE	R2
DT	0.903	1.508	1.222	0.947
RF	0.734	0.815	0.902	0.085
LR	0.693	0.785	0.885	0.021
GB	0.729	0.801	0.891	0.093
PH	0.708	0.778	0.882	0.002

From the results presented in Table 1 it can be concluded that the models do not perform well in predicting the defined objective, considering the metrics under comparison. Initially, the results seem to be quite satisfactory, since they present an MAE of approximately one person, however in percentage this value corresponds to a 100% error, since on average one patient is admitted per day in this specialty. Thus, the use of these variables added no predictive value since the use of time series also introduced no change.

After a detailed evaluation of each of the techniques, it was concluded that none of them presents results good enough to be implemented.

6. Conclusions

Focused on the CHTS' UCIP, this project defined a detailed characterization of it, including the prediction of future admissions based on exogenous variables for the period from 2017 to 2021. It was possible to analyse that, besides the fact that most of the admitted patients were male, there were a few more frequent origins, i.e., places prior to admission to this unit, such as the Emergency Department, the specialty of Surgery and Internal Medicine.

Five ML models (four regressions and a time series) were implemented, but no model obtained sufficiently accurate results, leading to the conclusion that the exogenous variables used had a low predictive value.

In terms of future work, it is expected the study of other exogenous variables to find out if they have greater predictive value. Is intended to develop a different solution to this problem, which may include the provision of daily admissions to this inpatient unit according to patients already admitted to the hospital, so it is possible to use, not only exogenous data, but also demographic and clinical data related to patients. Thus, despite the unsucces of the models developed, it is extremely important to look for other solutions, since the application of models with good results can bring several benefits to CHTS, as a better quality of care provided to the patient, given the possibility of having a more optimized bed management.

Acknowledgments

The work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: DSAIPA/DS/0084/2018.

References

- [1] H. C. Koh e G. Tan, «Data mining applications in healthcare.», *Journal of healthcare information management : JHIM*, vol. 19, n. 2, pp. 64–72, 2005, doi: 10.4314/ijonas.v5i1.49926.
- [2] Randhawa & Humayun, «Reasons of Overcrowding in Emergency Department», em *Journal of the Society of Obstetrics and Gynaecologists of Pakistan*, 1.^a ed., vol. 8, 2018.
- [3] R. Cipriano, «Grávida morre após transferência de hospital por falta de vaga», 29 de agosto de 2022.
- [4] J. Boyle et al., «Predicting emergency department admissions», *Emergency Medicine Journal*, vol. 29, n. 5, pp. 358–365, 2012, doi: 10.1136/emj.2010.103531.
- [5] A. Ribeiro, F. Portela, M. Santos, A. Abelha, J. Machado, e F. Rua, «Patients' Admissions in Intensive Care Units: A Clustering Overview», *Information*, vol. 8, n. 1, p. 23, fev. 2017, doi: 10.3390/info8010023.
- [6] H. M. Alshanbari, T. Mehmood, W. Sami, W. Alturaiki, M. A. Hamza, e B. Alosaimi, «Prediction and Classification of COVID-19 Admissions to Intensive Care Units (ICU) Using Weighted Radial Kernel SVM Coupled with Recursive Feature Elimination (RFE)», *Life*, vol. 12, n. 7, p. 1100, jul. 2022, doi: 10.3390/life12071100.
- [7] A. Vieira, «Modelação de Admissões e Internamentos na Urgência do Hospital de Braga», Universidade do Minho, 2016.
- [8] M. Silva, «Comparação de Modelos de Previsão da Chegada de Utentes ao Serviço de Urgência», Universidade do Minho, 2017.
- [9] V. Vijay, B. Kuechler, e S. Petter, «Design Science Research in Information Systems», n. 1, pp. 1–66, 2012, doi: 1756-0500-5-79 [pii]r10.1186/1756-0500-5-79.
- [10] C. Pete et al., «Crisp-Dm 1.0», *CRISP-DM Consortium*, p. 76, 2000.
- [11] G. James, D. Witten, T. Hastie, e R. Tibshirani, *Springer Texts in Statistics An Introduction to Statistical Learning - with Applications in R*. 2013.