UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial,
Aeroespacial i Audiovisual de Terrassa

Predicting hospital admissions to reduce crowding in the Emergency Departments of the Integral Healthcare System for Public Use in Catalonia

Document

MEMORY

Author

Joan Comalrena de Sobregrau Martínez

Academic supervisor

Jordi Cusidó Roura

Degree

MSc in Industrial Engineering

Call

April

END OF MASTER'S THESIS

Preface

The copyright of the master thesis rests with the author. The author is responsible for its contents. UPC is only responsible for the educational coaching and cannot be held liable for the content.

Abstract

Objective: This study analyzed data from Emergency Departments (EDs) from more than 60 different centers embedded in the Integral Healthcare System for Public Use in Catalonia (SISCAT) to predict hospital admissions based on information readily available at the moment of arrival to the ED. The predictive models might help reduce overcrowding at EDs and improve the service delivered to patients.

Method: A retrospective analysis was conducted using data from the SISCAT collected during the year 2018. Gradient boosting machine was used to train and test the predictive models in R, splitting the data in a 70/30 partition. Variable importance for each of the models was analyzed. Receiver Operating Characteristic (ROC) curves were created, and the Area Under the Curve (AUC) was obtained from each of them as a measure of predictive performance. The first part of the study targeted the obtention of models with high accuracy and AUC, while the second part targeted the obtention of models with a sensitivity > 0.975 and analyzed the possible benefits that could come from the application of such models.

Results: From the 3,189,204 ED visits included in the study, 11.02% ended in admission to the hospital. Gradient boosting machine proved to be a good method to predict for a binary outcome of either admission or discharge. The best performance for all the models was obtained at a 0.5 probability of admission threshold. The largest AUC was obtained for the complete dataset and yielded a result of 0.8938 with a 95% CI of 0.8929-0.8948. The best results for the sensitivity tests were obtained with the adults' dataset, with a model that gave a 0.4344 specificity and 0.5033 accuracy for a 0.975 sensitivity level.

Conclusion: This study reaffirms on the belief that gradient boosting machine is a powerful tool to use in binary outcome predictive models. It shows that data collected at the moment of arrival to the ED can be used to predict hospital admissions accurately, and that a model including data from a comprehensive hospital network has a better predictive performance when compared to a similar model developed with data from one unique health center only. It discusses the huge potential that the application of the models obtained could have in fighting crowding in EDs by allowing for an early start of the bed allocation process, making it possible to do all the required procedures for admission simultaneously to the patient being visited by the doctor, instead of doing it in a sequential manner after the visit, which unnecessarily crowds ED rooms and generates a non-optimal use of the available resources in EDs. The study also suggests the application of this predictive technique to develop models with proven high sensitivity to digitalize the patient-hospital relationship to allow for a first contact between both parties before the visit to the ED, which can potentially regulate the inflow of patients in this department and reduce ED overcrowding significantly.

CONTENTS

Introduction 5
Literature review 7
Research question and objectives 10
Research question 10
Main objective 11
Scope 11
Hypothesis 12
Method and material 12
Study setting 12
Feature extraction 12
Data preparation 13
Model fitting and validation 14
Results 15
Sample characteristics 15
Model performance 16
AUC testing 16
Sensitivity testing 19
Discussion 20
AUC testing 20
Triage method 21
Frequency of visits 21
Sensitivity analysis 22
Implications 23
Limitations 24
Economic Analysis 25
Conclusions 25
References 26

TABLE OF FIGURES

Figure 1. Sequential admission model 7

Figure 2. Simultaneous admission model 7

Figure 3. Too small learning rate λ[29] 9

Figure 4. Gradient Descent in GBM function[29] 9

Figure 5. Too large learning rate λ[29] 9

Figure 6. Patients included in the analysis 13

Figure 7. Accumulated number of visits per unique patient 15

Figure 8. Admission frequency per triage level 16

Figure 9. ROC curve for the complete dataset 17

Figure 10. Overfitting test. Optimal number of iterations 17

LIST OF TABLES

Table 1. Factors within hospitals affecting crowding[4] 6

Table 2. Variables included in the model 13

Table 3. Models' performance results 18

Table 4. Models' performance results (sensitivity > 0.975) 20

Table 5. Confusion matrix complete dataset 22

Introduction

During the last couple of years, the coronavirus pandemic has put healthcare systems from all around the globe under extreme stress. Most hospitals have seen their infrastructure and capacity to offer quality health care collapsing under an unexpected and enormous demand, and despite Covid-19 has obviously been the main actor in this crisis, the whole situation has evidenced some big inefficiencies of the system.

The healthcare service delivered in hospitals depends, to a large extent, on the efficiency in the execution of all those medical and non-medical processes aimed at providing, altogether, a quality service to the patient. These activities and relationships are often complex and multidisciplinary, making them a major focus of study as their optimization can result in improvements in many different areas simultaneously, such as the reduction in administrative costs, the optimization of hospital resources, the reduction in patient waiting times or a greater degree of service customization. This makes it possible to achieve, as a result, a higher quality service appreciated by both the institution and the user.

The Emergency Department (ED) is the service with the highest demand in a hospital, and this is continuously increasing, reaching to a critical point. The number of patients visiting the ED has been steadily rising worldwide for the last few decades, which has led to situations of overcrowding in emergency rooms all around the globe. Data from the Institute of Medicine in the United States already identified crowding as a critical threat for the quality of the service provided to patients back in 2006[1]. Similarly, the Spanish Ministry of Health released data regarding ED usage in 2010 in which it showed an increase of 23.2% of ED visits between 2001 and 2007[2], with an ED usage frequency rate notoriously higher than that of the UK or the US. The latest report showed a 9% increase in the number of visits attended in hospitals in Spain between 2014 and 2018, without any specific health reasons or population growth to account for that change.

ED crowding has been described by the American College of Emergency Physicians (ACEP) as "a situation in which the need for emergency services outstrips available resources in the ED. This situation occurs in hospital EDs when there are more patients than staffed ED treatment beds and wait times exceed a reasonable period."[3]

A health emergency, on the other hand, has been defined by the WHO as the unexpected appearance of a health issue of any type and cause, of varying degree of seriousness, that generates an imminent need of attention or treatment by a professional. Therefore, EDs need to offer a multidisciplinary assistance service, complying with functional, structural, and organizational requisites to guarantee safety, quality, and efficient conditions to attend any possible emergency.

However, crowding hinders the ability of EDs around the world to deliver such a service in optimal conditions. Studies show the direct correlation of ED crowding with increased ED waiting times, decreased patient satisfaction, inadequately treated pain, higher walkaway rates, and even higher mortality [4, 5, 6, 7]. And the cascade does not stop here. Hospital staff suffer from the consequences of overcrowding too, showing dissatisfaction, frustration, stress, and facing higher exposure to violence and physical aggression[8]. The optimization of this service and all the processes behind it must therefore be a priority in order to guarantee the continuity and improvement of its quality.

Factors	Influence
Increased ED demand	Extremely busy service results in overcrowding due to overcapacity
Increased hospital occupancy	Hinders bed allocation process
Increasing patient acuity	An aging population requires more workup and treatment
Patient self-referral	Patients who bypass the primary healthcare system overcrowd EDs
Inappropriate triage	Several hospitals implement non-approved triage systems that result in decreased efficiency and longer waiting times

Table 1. Factors within hospitals affecting crowding[4]

Reasons for the increase of this phenomenon are varied (*Table 1*). Some experts point to patients being increasingly demanding in the immediacy to receive health checks for non-urgent sufferings. However, professionals in the field also identify problems in the structure, that is not always able to solve problems quickly enough in the primary attention system, which is marked by long waiting times and is therefore being abandoned by users.

One of the processes that has greatest impact on emergency room congestion is the admission of patients from the ED into the hospital, as a result of the required logistics for patient management and bed allocation.

The impact of an inefficient patient admission process in the ED is of high relevance. While only around a 10% of the visits in the ED department end in an admission to the hospital, EDs are still the largest source of hospital admissions. Lengthy boarding times of patients that are waiting to be assigned a bed at the hospital use precious resources such as time, space, and medical attention - that should instead be used in other ED tasks- and contribute to the overcrowding of the service[9]. The accumulation of patients in the ED generates a difficult to dissolve bottleneck in the admissions process and hinders the capacity of the ED to respond to the continuously arriving demand of new admissions.

Various solutions to solve this situation have been proposed during the last few years, the most significant of which being a simplification of the admission procedure[10], the inclusion of more doctors in the admission process of both the hospitalization plant and the emergency department[11,12], an optimized process of early hospital discharges[13], or the creation of an ED dependent unit for short-length stays[14], but none of them seem to be easily actionable without large economic efforts, or sustainable in a situation where ED visits will continue to increase.

In contrast, monitorization of the admission process and anticipation of hospital admissions can potentially help optimize the use and allocation of resources in a sustainable way, thus improving the quality of emergency care and user satisfaction. Data collected in hospital databases and shared through Health Information Exchange (HIE) platforms can be used to make early predictions of inpatient admission and help in the implementation of actionable measures. In the currently used system, the process of inpatient admission to the hospital from the ED starts after the visit is completed. That means that after the visit, the patient must wait in ED facilities -crowding them unnecessarily- while the administrative staff processes their admission and allocates them a bed in the hospital. With an early prediction of admission, for example at the moment of arrival of the patient in the ED, the administrative staff would be able to carry out this process while the patient goes through the ED visit in a simultaneous rather than sequential way. By doing so, if the patient had to be indeed admitted into the hospital after the visit, the admission process would have already been completed and there would be no extra waiting time nor unnecessary crowding, as portrayed in *Figures 1 & 2*.

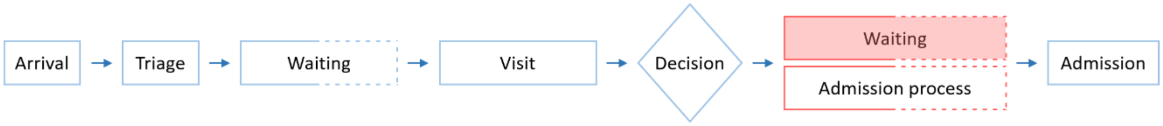

Figure 1. Sequential admission model

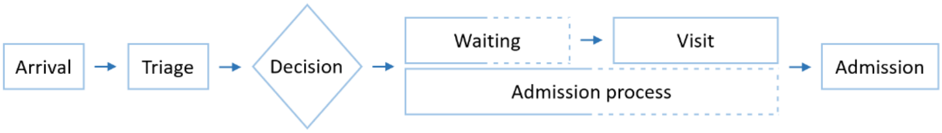

Figure 2. Simultaneous admission model

Although some studies have demonstrated that data collected during the patient visit can slightly improve predictive performance as compared to that of a model to predict admissions at time of arrival, the models developed in this study aimed at predicting admissions with information readily available at moment of triage, or what is the same, right after arrival of the patient in the ED. This was done this way simply because the sooner the admission process can be started, the higher the ability to reduce ED crowding can potentially be, which is, in the end, the final objective of the development of a predictive model of hospital admissions.

Literature review

Prediction models in healthcare seek to increase logistical efficiency, enhance resource allocation, reduce boarding times, and improve patient care.

During the last fifteen years, researchers have put attention in data collected early in ED encounters to help existing triage systems more quickly identify and prioritize patients with critical conditions from the volumes of those with less urgent needs to tackle and possibly alleviate ED overcrowding.

Triage is the first stage of an ED visit. Its objective is to assess in a regulated, validated and reproducible way, and as accurately as possible, the level of urgency of a visit to organize patients in recognizable groups to prioritize the sickest ones[15]. There are many different triage systems. The most widely accepted ones are the Australian Triage Scale (ATS)[16], the Canadian Triage Acuity Scale (CTAS)[17], and the Manchester Triage System (MTS)[18]. However, there are other approved systems such as the Andorran Triage Model (MAT)[19], which is used in almost all the hospitals in the Catalan healthcare system[2].

Researchers have been implementing machine learning algorithms -in different forms- to extract valuable information from the huge amounts of data existing in hospital -and ED- databases. These algorithms are capable of understanding patterns in data and building corresponding models that use the identified patterns to classify new observations.

Some basic studies have researched in the topic with the development of models to predict future ED visits, thus facilitating provision and preparation of ED staff to avoid overcrowding. Penades and

Ros (2015) used ARIMA and Holt-Winters models to demonstrate that such predictions were possible with mean errors of 2.5% and 3.84% respectively when predicting visits at a monthly level[20].

More complex models have attempted to predict, not the number of visits, but the percentage of these that will turn into an admission to the hospital. The first studies were centered in trauma level I centers, in which the reasons for visiting are not as varied as in a general hospital ED, and severity of disease (or injury in this case) can be more objectively assessed at triage. Probabilistic systems in the form of Bayesian network for early prediction of admission in ED in these centers, with data available early in the ED visit, proved to be useful for predicting patient admissions already in 2005 (AUC = 0.894)[21]. The same researchers concluded in 2006 that an Artificial Neural Networks (ANN) could also be used to predict hospital admissions in pediatric ED encounters, again in trauma level I centers, showing an almost equal performance (AUC = 0.897)[22].

More recently, logistic regression and Naive Bayes analysis have been popularly used to make predictions in the healthcare environment. Savage et. al. (2017) applied logistic models on triage administrative data to estimate admissions in the ED department (AUC = 0.78) and use it to predict, in turn, hourly bed requirements. The logistic regression models obtained had a sensitivity of 23% and a specificity of 97%. Although these results were by themselves satisfactory for particular individual admission predictions, the hourly pooled probabilities of bed requirements showed better results, falling close to historical demand data[23].

Barak-Corren, Fine, & Reis (2017) looked for improved accuracy (percentage of predictions, of both admissions and discharges, that conform to the real observed value) on the combination of different analytical tools with the logistic regression approach. Applying a logistic regression model on results generated by a Naive Bayes classifier from data collected within the first 30 minutes from arrival yielded good results in their study, identifying more than 73% of admissions with a 90% specificity and over 35% with 99.5% specificity, or what is the same, a false-positive rate of 0.5% (AUC = 0.91)[9]. They also applied the method to predict admissions at the moment of arrival (after 0 minutes), successfully identifying 50.6% of the hospitalizations with a 10% false-positive rate and obtaining an overall AUC = 0.79. (The specificity in these models to be understood as the capacity of the algorithm to correctly predict the discharges, i.e. a specificity of 90% means 90 out of 100 discharges were predicted as such with the model, while 10 out of 100 were wrongly predicted as admissions).

Other researchers have tried manipulating the data slightly to obtain models that would show better performance measures and higher accuracy. Lucke et. al (2018) showed that it was possible to increase prediction accuracy of admission at time of arrival for patients visiting the ED with a logistic regression model by dividing the observations in two groups, one containing those related to patients under 70 years old (AUC = 0.86), and a second one including the rest, i.e. those containing observations of patients above 70 years old (AUC = 0.77)[24].

The random forest technique has also been used in this field to develop an e-triage model to predict likelihood of acute outcomes -that may lead to admission-. Levin et. al (2018) obtained models with AUC ranging from 0.73 to 0.92 for different datasets[25]. Although the objective of their study was more focused on the demonstration that the current triage systems could be improved through the use of information extracted from patient data upon arrival at the ED, it also showed the potential of these models to predict hospital admission with significant accuracy.

Finally, a study from 2018 set the objective to determine which, among the most popular predictive techniques, could provide the best performance. Training models on triage information yielded a test AUC of 0.87 for all logistic regression, gradient boosting, and deep neural networks (DNN) models. However, the study went further on to proof that combining triage information with patient history could significantly improve predictive performance for all three methods, achieving an AUC of 0.91 for the logistic regression model, and of 0.92 for the gradient boosting and the DNN models. Models trained on patient history information exclusively did not yield better results than those trained with triage data alone[26].

Despite knowing that gradient boosting is a powerful tool for predictive modeling, machine-learning and data-mining, and one of the best algorithms in classification tasks, providing higher accuracy and better results than other conventional single strong models and a better predictive performance than logistic regression, no studies had used it before with the objective of predicting hospital admissions in EDs at the moment of triage. This study will test its true potential.

Gradient boosting is a machine learning algorithm that builds an ensemble of weak trees in a sequential fashion, with each tree being trained with respect to the previous and reducing the marginal error of the whole ensemble learnt so far[27]. Subsequent trees therefore help classify observations that are not well classified by previous trees. Later, weak trees are combined in a gradual and additive manner into a powerful and strong model with high predictability, hard to beat with other algorithms[28]. The theory behind this process is to train the new base-learners to be maximally correlated with the negative gradient of the loss function associated with the whole ensemble until the gradient descent is zero (see *Figure 3*).

For datasets with a categorical response $y \in \{0, 1\}$, the two most used loss functions are the Binomial and the Adaboost loss function, which are more generally referred to as the Bernoulli loss functions.

Given that gradient boosting machines learn sequentially from previous weak models, it is of high relevance to define an adequate shrinkage value. This value can help regularize and control model complexity by potentially reducing the impact of unstable regression coefficients reducing the size of incremental steps in the model learning process. The main principle applied is that taking many small steps in

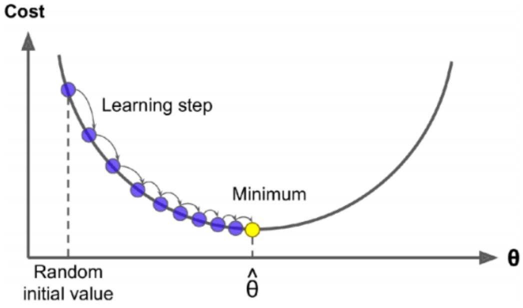

Figure 4. Gradient Descent in GBM function[29]

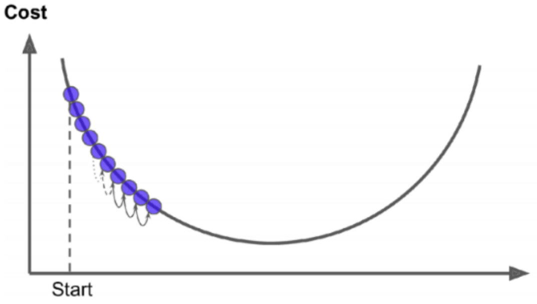

Figure 3. Too small learning rate λ[29]

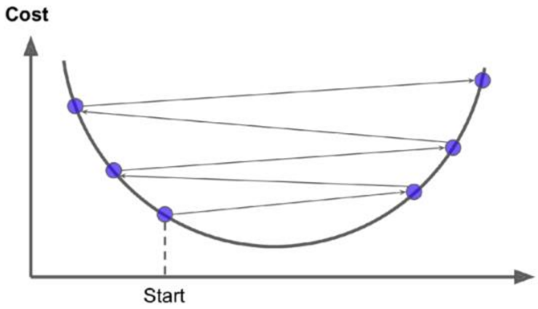

Figure 5. Too large learning rate λ[29]

improving the model is better than taking fewer large steps. The shrinkage parameter is defined as $\lambda \in (0, 1]$[30]. Although the number of weak trees to train is defined by the researcher, for small values of λ the gradient boosting algorithm will require many trees (>1,000) to arrive to a satisfactory result (see *Figures 4 & 5*), which can be computationally expensive and timely consuming.

However, there is a cost to reducing shrinkage too much, known as overfitting. This is defined as an excessive improvement of the model to the training dataset, learning and adapting too much on its particularities and resulting in a decrease in the performance on the test dataset. This can be solved with the early stopping technique, which allows the algorithm to stop the number of iterations before the initially pre-specified one if it detects overfitting, or what is the same, a decrease in the predicting power of observations in the test set. This optimal number of boosts is therefore dependent on the shrinking parameter λ. To deal with the trade-off existing between the learning rate and number of boosts is usually approached using a cross-validation procedure, which allows for testing the model on withheld portions of data, while still using all of the data at the processing stages. To do so, a cross-validating parameter *k* is defined and used to partition the data into *k* disjoint non-overlapping subsets, each of which will be used as a validation set for the GBM model fitted with the rest of subsets. Only after doing the process for all the subsets will the validation performance from each of the folds be aggregated to serve as estimate of model generalization on the validation set[27].

One of the disadvantages of gradient boosting machines -apart from overfitting- is the comparably low interpretability of the results obtained. There are two main tools to address this issue. The first one is the relative variable influence analysis, which is a common tool used in many cases to base the feature selection on, although it does not provide any specific explanation on how each variable actually affects variation in the result. The second tool are partial dependence plots, which are commonly used to visualize the effect of one selected variable on the response while controlling for the effect of all other explanatory variables. These graphs may be less insightful when interactions between variables have significant impact on result, but they have been proven to provide a solid basis for interpretation of the models[31].

The most valuable feature of GBMs is that they are highly flexible. As mentioned, parameters such as number of trees, depth of trees, learning rate and subsampling can be modified to tune the training model and improve its efficiency while enhancing overall performance[32].

Research question and objectives

Research question

Based on prior research, it is safe to assume that patient data collected at the moment of triage can be used to predict hospital admissions with good accuracy. However, all the studies mentioned have developed models based on hospital-specific data, which makes them little scalable for wider healthcare networks. In this study it will be attempted to develop a model to predict hospital admissions in Catalonia with data from the Integral Healthcare System for Public Use in Catalonia comprising more than 60 different health centers[33].

The main research question is to determine whether it is possible or not to predict hospital admissions based on patient data collected at the moment of arrival to the ED, from a large number of different health centers, with an AUC over 0.87 -which is the best result observed in the literature for a dataset with similar variables using a logistic regression model[26]-, using the gradient boosting machine learning algorithm. Determine whether GBM can beat the performance previously observed in logistic regression models.

Main objective

The main objective of the paper is to present a predictive model of hospital admissions, using data from a large healthcare system, with special focus on the accuracy of such a model but also looking at its sensitivity -for the applications that a model with very few false-negatives could have in the healthcare industry-. This would allow for a very much needed early prediction of admissions in EDs that would help address overcrowding through a more efficient bed allocation process and therefore reduce boarding times and increase overall patient satisfaction.

As previously defined in the literature, the accuracy is the ability of the model to correctly -or accurately- predict an outcome, either admission or discharge. Sensitivity, on the other hand, is the ability of the model to correctly classify observed admissions as such and can be obtained dividing the number of true positives by the sum of true positives plus false negative, or what is the same, the total number of observed admissions. An effort to increase sensitivity can yield lower scores of specificity, as a result of an increase in false positives.

Scope

Assess whether the currently used triage system (MAT) can successfully classify patients based on the urgency of their sufferings and see if the values assigned correlate with the admission result for every individual observation.

Analyze the relative importance of the variables used for the development of the predictive model. Rank the variables based on information gain.

Look for inefficiencies in the structure of the emergency service based on patterns and outlier values in the data that may call for a restructuring of the system in the way patients are admitted, triaged, and treated. Detect, among those outliers, symptoms of emergency health service abuse.

Present data to support the development of a software for patient remote-use previous to the ED visit, to reduce the number of visits received in EDs in the Integral Healthcare System for Public Use in Catalonia (SISCAT).

Establish the bases for future research studies at regional or national level including patient history data, which has the potential to increase performance of the models by more than 5 percentual points as shown by Hong WS, et. al (2018)[26].

Hypothesis

H_1: It is possible to develop a statistical model using Gradient Boosting that can predict, with an AUC higher than 0.87 (the best result observed in the literature for a dataset with similar variables, using a logistic regression model[26]), whether a visit to the ED will result in a patient admission to the hospital based on data available at the moment of arrival at the ED.

H_2: Splitting the dataset into smaller subsets based on the age of the patient can increase the model's predictive power. Based on Lucke, J.A. et. al (2018) results in the field of hospital admissions prediction, these splits will be done at 18 and 70 years of age.

H_3: Principal diagnosis information (based on the Clinical Classifications Software of the US government)[34] is the variable with the highest relative importance in the models.

H_4: Triage level is negatively correlated with admission probability. The lower the triage level, which entails higher priority, the higher the admission probability.

H_5: Higher visit frequency does not lead to higher admission rate. This non-existent correlation can be used to detect ED service abuse and inefficiencies in the system.

Method and material

Study setting

A retrospective analysis of all visits to EDs of hospitals included in the Integral Healthcare System for Public Use in Catalonia from January 1, 2018 to December 31, 2018 was conducted, including 3,189,204 observations.

The data was provided by the Catalan Government through the Catalan Healthcare System (Servei Català de Salut, CatSalut) at date May 21, 2020. It included all the observations that had been correctly identified with the personal ID of the patient. In this pre-processing stage carried out by the governmental entity 127,806 observations were left out due to missing or erroneous identification numbers (ID), accounting for a 3.6% of the total emergency visits. The remaining observations, correctly identified, recorded disposition of either admission or discharge after the visit at the ED. The data was provided in .dat format, which could be directly imported to Rstudio for processing.

Feature extraction

For each observation, the following fields were included: identifier, age, gender, main diagnosis of the urgency based on the CCS system, triage level according to the MAT system, and admission result in binary format. Age information was either obtained at triage or available from the Electronic Health Records; for patients under 1 year of age, an extra variable containing the age in days was included. Different visits with the same identifier were classified as different observations.

Data preparation

Looking to increase predictive performance of the model, some variables were added, and some others modified. First, however, the dataset was cleansed to eliminate all those observations that were incomplete, which accounted for 189,337 observations (5.51%). Following this first cleansing, a quick analysis was performed to identify variables that might have abnormal values; 34,725 (1.01%) observations were eliminated for containing non-existent triage values, i.e. >5. In the category of symptoms (CCS), some observations had been classified as 'incomplete' and were eliminated from the dataset as well; that accounted for 20,864 observations and a 0.61% of the initial total.

After this pre-processing stage of the data performed by both the government and the researcher, the number of observations for analysis had dropped from 3,561,936 to 3,189,204. (See *Figure 6*).

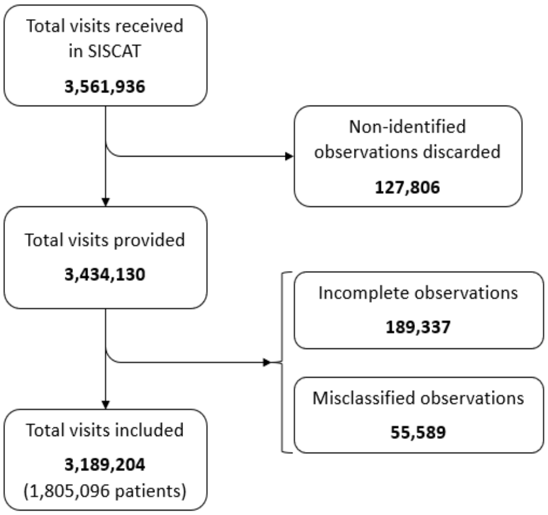

Figure 6. Patients included in the analysis

Observing that it was recurrent to have multiple visits for one patient in the 2018 exercise, an extra variable was created to show the accumulated number of visits for each patient during the year under study. The same procedure was used to see the absolute frequencies for the symptoms (CCS) variable, and a feature selection procedure was conducted to reduce the number of levels with very little frequency. In this case, it was decided to assign the label 'Other' to all these observations whose CCS value appeared fewer than 30 times throughout the dataset, which would imply an average of less than one case every ten days in the whole Integral Healthcare System for Public Use in Catalonia, on average.

Table 2 shows the number of distinct outomces observed for each of the variables in the dataset.

Variable	Distinct outcomes
ID	1,805,096
Accumulated visits	77
Age	113
Age Days	365
Gender	2
CCS	259
CCS frequency	255
Triage	5
Admission	2

Table 2. Variables included in the model

Some of these variables were classified as numerical by default although they were categorical. For that, Gender, CCS and Triage were factorized before moving on with the data analysis.

To test Lucke, J.A. et. al (2018) conclusion stating that accuracy of the prediction at time of arrival for patients visiting the ED could be improved by dividing the observations in two subsets, one containing all those related to patients under 70 years of age and another containing all those above 70, a division of the main dataset was done for this study as well. To further contrast this enhanced predictive value, another division was created with the 18 years of age threshold in order to test the performance of a model dedicated to pediatrics against that of a model for adults. This resulted in 5 distinct datasets: the complete dataset, the pediatrics dataset [0-18) years, the adults dataset [18, 115] years, the adults under 70 dataset [18, 70) years, and the adults over 70 dataset [70, 115] years. For all the datasets including adults exclusively, the variable Age Days was eliminated.

Model fitting and validation

The models were trained on each of the five datasets described above using Gradient Boosting Machine in R, using the 'gbm', 'caret', 'ROCR' and 'pROC' packages. Each of these datasets was divided into training set and test set, randomly sorting 70% of the of the observations to the former and the remaining 30% to the latter. The training set was used to build the model for the gradient boosting machine, analyze the variable importance in predicting the outcome, and test for overfitting. The test set was used to evaluate the predictive performance of the model trained with a confusion matrix from which to extract the performance parameters (accuracy, sensitivity, specificity, PPV, NPV) and to create the ROC curve, therefore obtaining the AUC for each of the models, with a 95% confidence interval obtained with the DeLong method. No further data pre-processing was performed, given that Gradient Boosting does not require it.

The Gradient Boosting models were developed with a Bernoulli distribution function, since it was considered the best option to predict for the binary response. 1,000 trees were used to train all the models, with an interaction depth of 3 levels and a shrinkage equal to 0.01. To be sure that the trained models were not overfitted, a tuning process was performed to all of them using the exact same parameters and adding a 2-fold cross validation.

Variable importance is an indicator of the information gain that a given variable provides in a split. The rank for variable importance was obtained for each of the models and later used to create models including only the 3 most influential variables and see if it was possible to obtain the same levels of predictive performance.

Predictions for the accuracy testing were done using 1,000 trees as well and setting a threshold for prediction to 0.5; that is, if a prediction yielded a result with more than 50% probability of being an admission, it would be classified as an admission. Confusion matrixes were obtained from the predicted results. Information about performance was also obtained from the Receiver Operating Characteristic curve for each model, where the true positive rate was plotted against the false positive rate. Moreover, these curves allowed for the obtention of the Area Under the Curve, which was provided with a 95% confidence interval following the method described by DeLong[35].

With the same models used for the first part of the research, that targeted high accuracy results, the second part of the study was approached. In this, the objective was to obtain high sensitivity levels (>0.975) while maintaining the specificity over 0.33. To do so, and with the aim of predicting as many admissions as possible, the probability of admission required to classify an observation as such was lowered substantially, and modified in a range from 0.05 to 0.01. This procedure was conducted for each of the models until the required sensitivity level was obtained, and the specificity level was checked after reaching the critical point. This models presented a trade-off between the ability to detect 97.5% of all the observed admissions, and the ability to correctly predict as many outcomes as possible. As will be discussed below, targetting high sensitivity levels made the number of false positives increase substantially. The models obtained would therefore not be very adequate for implementation in the ED, for the staff would anticipate more admissions than really observed, and a lot of resources would be wasted in preparing for these false-positives. However, these models can have a very high potential of implementation, not for hospital use but for patient use, to allow for a first contact between patient and ED, and regulate the number of patients self-referring themselves to the ED for non-urgent causes. A more in-depth explanation has been developed in the discussion.

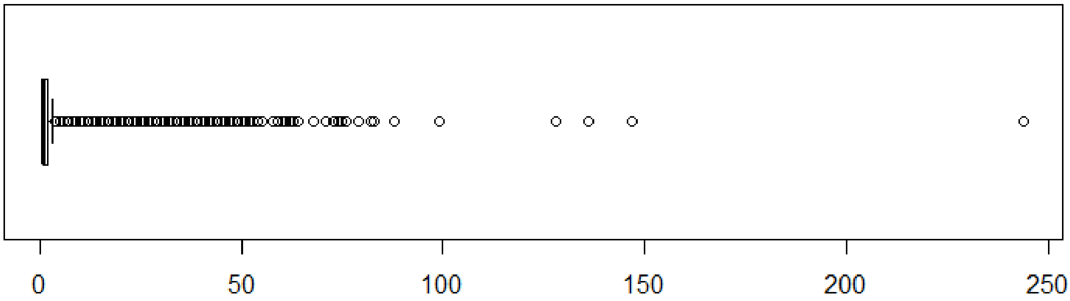

Figure 7. Accumulated number of visits per unique patient

Results

Sample characteristics

During the year 2018, the Emergency Departments in the Integral Healthcare System for Public Use in Catalonia received a total of 3,561,936 individual visits, 3,434,130 of which were included in the dataset provided by the Catalan Healthcare System to perform the study. After cleansing the data, 3,189,204 observations that had been correctly identified and classified were included. These accounted for 1,805,096 unique patients.

The accumulated number of visits per patient ranged from 1 to 244, with a median of 1, and a mean equal to 1.767. A box plot of the absolute frequencies of visits per patient can be found in *Figure 7*, in which the outlier values regarding number of visits are clearly visible.

The overall admission rate was 11.018%, with 351,391 admissions by 275,875 unique patients, and 2,837,813 discharges by 1,690,092 different patients.

Looking for an indicator of admission in the frequency of visits, a correlation analysis was performed using the *Pearson* correlation method. This yielded a result of 0.01, an extremely week correlation that indicates both variables are hardly related. Further, a p-value < 0.001 was obtained, giving undeniable significance to the test result. To observe the behavior of these two variables against each other, a graph was created (see *Appendix I*).

The gender distribution of the sample was 46.44% male and 53.56% female, with a rate of admission slightly higher for the former, 11.41% against 10.68%, that proved to be statistically significant with an alpha equal to 0.001 (p-value < 0.001). The average age of the patients included in the study was 43.08 years old, being slightly higher for females (44.90), than for males (41.91). Charts for the gender specific age distribution can be found in the *Appendix II*. The average age for those visits that ended in an admission in the hospital was 59.57 years old, while that of those visits that ended in discharge was 41.04 years old, a more than 18 points difference that was clearly significant with a p-value < 0.001 for an alpha equal to 0.001 (see *Appendix II* for graphic visualization).

Regarding the triage level frequencies observed and knowing that 1 accounts for the most urgent and priority cases and 5 for the least urgent ones, it is worth noticing the statistically significant higher severity (p-value < 0.001, alpha = 0.001) of the classification of patients admitted in front of patients discharged; the former had an average triage level of 3.03, while the latter had a 3.74 average triage level.

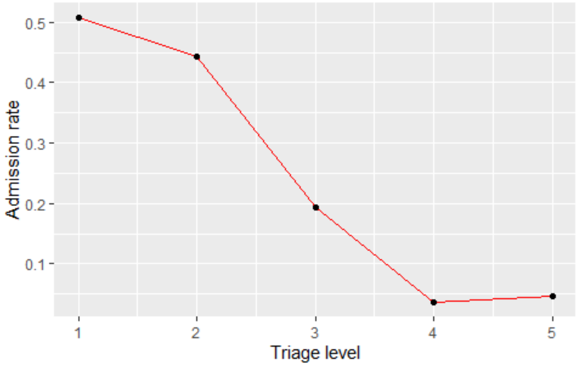

Figure 8. Admission frequency per triage level

Figure 8 shows the admission rates for all five triage levels. Looking at absolute figures it is easy to see that very few cases are classified as extremely acute (level 1), and that the big bulk of patients is classified in either level 3 or 4. The number of patients for each triage level was the following: level one, 5,774 observations, level two, 161,714 observations, level three, 1,048,227 observations, level four, 1,650,429 observations, and finally level five, 323,060 observations.

All these previous statistical analyses were performed with the student's t-tests method to compare means from different samples with two-sided critical area (two-tail t-test).

The most predominant symptoms classified at arrival of the patient in the ED were: spondylosis, intervertebral disk disorder, and other back complaints (4.43%, 141,357 observations), followed by superficial wound & concussion (4.23%, 134,970 observations), abdominal pain (4.12%, 131,261 observations), non-classified codes for causes that were unclear, not included in the CCS or that the patient could not describe (3.96%, 126,190 observations), and other respiratory infections in the upper tract (3.65%, 116,420 observations). On the other hand, the top five symptoms leading to a higher percentage of admitted patients were polyhydramnios and other disorders in the amniotic cavity with a 92.0% of the visits ending in admission (6145 observations), followed by appendicitis and other appendicular affections, with an admission rate of 88.2% (5633 observations), non-diabetic pancreatic disorders with 87.1% (3864 observations), umbilical cord complications with 87.0% (31 observations), and finally, femoral neck fracture showed an 86.9% admission rate (7,073 observations). See *Appendix III* for more information on most frequently observed symptoms.

Model performance

AUC testing

Complete dataset

The first model analyzed was developed with the complete dataset, including all the observations deemed valuable (see *Figure 6* in Data preparation). The accuracy of the model was 0.9113, exceptionally high when compared to the results obtained by other studies. The AUC for this model obtained from the ROC curve was 0.8938 with a 95% CI of 0.8929-0.8948, also a better result than any observed -at time of arrival to the ED- in any previous study.

Figure 9 shows the ROC curve obtained for this first model, varying the discrimination threshold from 1 (bottom-left corner) to 0 (upper-right corner). As always sought in ROC curves, the objective is to have the curve as close to the upper left corner as possible, which implies a high true positive rate (sensitivity), with a low false positive rate (1 – specificity). All the ROC curves for the following models had the same shape, so they have not been included in the report to avoid repetition.

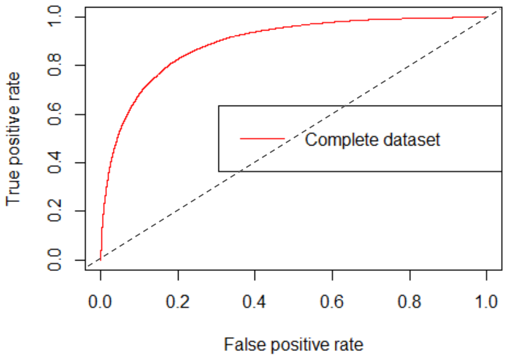

Figure 9. ROC curve for the complete dataset

The importance of the variables in the model was obtained with the summary function of the 'gbm' package and it showed that the symptom -classified with the CCS method- was the variable providing the higher information gain at every tree split (76.98%), followed by the triage level (16.68%) and age (6.08%). The four remaining variables (accumulated number of visits during the year, days of age for babies, gender, and frequency of appearance for CCS symptoms) had an almost non-significant contribution to the model adding up to 0.25% of the explanation of it, with frequency of CCS symptoms showing a null relative influence. This information was used to create a model including only the first three variables and to run the analysis again. This yielded results similar to these of the complete model, with an accuracy of 0.9112 and an AUC of 0.8936 (95% CI 0.8927-0.8946). More information about variable importance can be found below in *Appendix IV*.

The initial model was tested for overfitting, but the results clearly showed that using 3 levels of depth for the trained trees and a 0.01 shrinkage parameter, 1,000 trees were an optimal amount and not large enough to run into overfitting. *Figure 10* shows the error on both the train (black line) and test (green line) subsets. It is clear that the performance on the test set is not being jeopardized by an overfitting to the training set characteristics. The dotted blue line shows the optimal number of iterations (trees) given the input parameters.

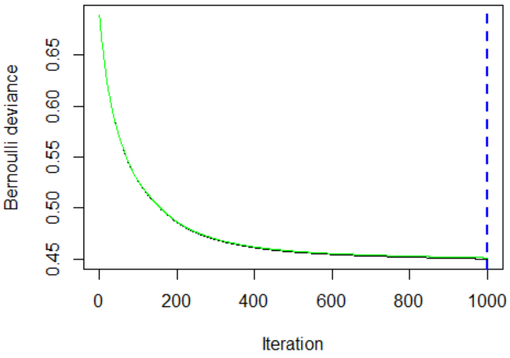

Figure 10. Overfitting test. Optimal number of iterations

Pediatrics dataset

This subset included all the observations of patients from 0 to 17 years of age inclusive, which added up to 687,288 observations of 381,470 different patients. This model provided an accuracy of 0.9577, with an AUC of 0.8703 (95% CI 0.8667-0.8739). Note that this model has higher accuracy but lower AUC than the complete dataset; this is an indicator of this model being very good at identifying discharges but not so good at identifying admissions, which is probably a consequence of a very imbalanced set in terms of the ratio of admissions/discharges. The importance of the variables in this model differed from that observed in the complete dataset as well, with symptoms and triage still occupying the first and second places (69.13% and 25.97% respectively) but with days of age in the third position (3.02%) right in front of age in years (1.79%).

Adults' dataset

This subset included 2,501,916 observations from 1,425,606 unique patients ranging from 18 to 115 years of age. The predictive model obtained had an accuracy of 0.8985 and an AUC of 0.89112, with a 95% CI of 0.8901-0.8922. Variable importance for this model showed the same rank as for the complete dataset with slightly different contributions; 79.95% for CCS symptoms, 15.25% for triage and 4.54% for age.

Young adults' dataset

The young vs. old adults' division was performed to test the results obtained by Lucke, J.A. et. al (2018).

This one included 1,819,221 observations from 1,070,125 different patients and provided an accuracy of 0.9292. The AUC for this subset was 0.89114 with 95% CI of 0.8896-0.8926. The most important variable was, again, CCS symptoms (81.29%), followed by triage (16.26%) and age (1.81%).

Old adults' dataset

The second adults' subset included 682,695 observations from 357,913 unique patients. The predictive model developed provided an accuracy of 0.8182 with an AUC of 0.8514 (95% CI 0.8496-0.8533), clearly ranking last in performance among all subsets. Variable importance consistently showed CCS symptoms as the most informative variable, with triage and age occupying the second and third places (85.64%, 13.36% and 0.765% respectively).

Additional information about variable importance and the performance obtained for each of the models can be found in *Appendix IV*.

Summary of results

	Complete DS	Pediatrics DS	Adults DS	Young adults DS	Old adults DS
AUC	0.8938	0.8703	0.89112	0.89114	0.8514
Accuracy	0.9113	0.9577	0.8985	0.9292	0.8182
Specificity	0.9777	0.9959	0.9709	0.9875	0.9181
Sensitivity	0.3746	0.1577	0.4019	0.3116	0.5007
PPV	0.9267	0.9611	0.9175	0.9383	0.8538
NPV	0.6752	0.6478	0.6687	0.7015	0.6581

Table 3. Models' performance results

Sensitivity testing

As a reminder, in this part of the test the admission prediction threshold was moved down from 0.5 to as much as needed to obtain a sensitivity >0.975.

Complete dataset

For the complete dataset, a sensitivity of 0.975 was obtained with a probability of admission prediction threshold of 0.022. At this point, the specificity of the predictive model had dropped below 0.5 -to 0.4262-. The overall accuracy obtained for that model was 0.4867. From the 105,346 admissions in the test subset only 2,574 were missed by the model, which came at a cost of correctly predicting just 362,891 discharges out of 851,415.

Pediatrics dataset

The model for the pediatrics dataset performed the worst, only reaching the 0.975 sensitivity goal at 0.01 probability of admission prediction threshold. By that point, specificity was at 0.3540 and accuracy at 0.3823. This might be explained by the low number of admissions in the pediatrics test subset and for the struggle that it can be to make an accurate guess of the symptoms responsible for babies' visits.

Adults' dataset

As with the previous models, the first iteration was done with a probability of admission prediction threshold equal to 0.05. This had to be lowered down to 0.024 to obtain a sensitivity of 0.975, which was obtained together with a specificity of 0.4344 and an accuracy of 0.5033. For this model, the number of false negatives was as low as 2,317 out of the 95,590 observed hospital admissions, and only 284,521 discharges were detected for 370,463 that were missed.

Young adults' dataset

In comparison to the adults' model, and following patterns observed in the pediatrics' analysis, this first adults' subset required the probability of admission prediction threshold to be lowered down to 0.017 (vs. 0.024 for the complete adults' subset) to obtain the desired 0.975 sensitivity. The specificity and accuracy of the model at that point were 0.4163 and 0.4645 respectively.

Old adults' dataset

Maybe not as surprisingly as it may seem at first, with a probability of admission prediction threshold of 0.05 this subset had a sensitivity higher than the desired 0.975, and it allowed to increase this threshold up to 0.055. The specificity obtained was in this case of 0.3614, while the accuracy remained slightly over 0.50.

The confusion matrixes with the absolute values of accurate predictions and false positives and false negatives can be found in *Appendix V*.

Summary of results

	Complete DS	Pediatrics SS	Adults SS	Young adults SS	Old adults SS
Sensitivity	0.9756	0.9754	0.9760	0.9752	0.9755
Specificity	0.4262	0.3540	0.4344	0.4163	0.3614
Accuracy	0.4867	0.3823	0.5033	0.4645	0.5085

Table 4. Models' performance results (sensitivity > 0.975)

Discussion

This study has shown that it is possible to predict hospital admissions in Emergency Departments in an integral healthcare network of more than 60 hospitals in Catalonia, using only information collected at time of arrival to the ED. This can have big implications for hospitals trying to reduce ED overcrowding and speed up the bed allocation process for patients from the ED, which is one of the most relevant bottlenecks in the inpatient admission process and a main cause of ED crowding. Furthermore, the study has proved it is also possible to obtain predictive models with remarkable sensitivity levels without sacrificing specificity excessively, which might have big application potential to reduce ED overcrowding, as will be discussed below.

AUC testing

The first thing that was done was to analyze the predictive power of a model for the whole population sample, from 0 to 115 years of age, with the gradient boosting algorithm. Based on previous literature, such a model would be satisfactory if it were able to provide an AUC comprised between 0.8 and 0.87. Savage et. al. (2017) obtained a 0.78 AUC[23], Barak-Corren, Fine, & Reis (2017) obtained a 0.79 AUC[9], Lucke et. al (2018) obtained with a u70-year-old patients' dataset a 0.86 AUC[24], Hong WS, Haimovich AD, Taylor RA (2018)[26] obtained an AUC of 0.87. However, the goal was set to improve previous research by using a more comprehensive dataset and the GBM method to obtain an AUC of 0.87 or higher (as stated in H_1). This goal was achieved with the complete model, which yielded an AUC of 0.8938 (95% CI 0.8929-0.8948), beating the best performance found in the literature for the prediction of admissions at time of arrival to the ED, which was obtained by Leegon, J., Jones, I., Lanaghan, K., & Aronsky, D. (2005)[21] in a trauma level I center -a sample clearly not as complex as a complete integrated healthcare system-. This performance makes it possible to accept H_1. Since this value was the highest AUC result observed in any previous research, it can be taken as an indicator that using a comprehensive dataset can indeed improve predictive performance for ED admissions. However, it is difficult to discriminate whether this great result was obtained thanks to the dataset or thanks to the method, so this will be included as a limitation of this study and proposed for future research. As shown by Hong WS, et. al. [26], patient history data could further improve the predictive performance of a model such as the one developed in this study (these researchers managed to improve predictive performance by a 5% with this additional information). In an attempt to integrate this data in the study, a formal request was made to the Catalan Government to obtain patients' medical histories but given the Covid-19 situation and the overload of work in the healthcare network in Spain, the governmental entity deemed it would entail too much data preparation work and that could not be assumed by the hospital IT departments, so that is an issue that should be targeted in future research.

The second possibility to improve predictive performance was to divide the dataset into smaller subsets based on patient age, which was hypothesized as possible in H_2 following the results obtained by Lucke et. al (2018)[24]. This was done in two steps, first using the 18 year old threshold to separate pediatrics visits from adults visits, and second by dividing the adults subset at the 70 year of age threshold, which for the already mention researcher provided a predictive improvement. This hypothesis was however not accepted based on the results obtained. In the first division, the averaged AUC between the pediatrics and the adults' datasets obtained was 0.8866 (95% CI 0.8851-0.8883), which was lower and fell outside of the 95% CI for the AUC results for the complete dataset. The averaged AUC after the division between young adults and old adults was 0.8803 (95% CI 0.8787-0.8819) which was not able to beat the previous results obtained for the complete adults' dataset (AUC = 0.8911, 95% CI 0.8901-0.8922). These counterintuitive results may have been caused by the high influence of wrongly classified observations, especially in the very old and very young age ranges, that ended up showing a different result than the one expected and biased the predictive ability of the model to correctly classify future observations. However, it should not be considered a big issue since the differences between AUCs of different models were of the range of 0.1 difference.

The results obtained in the study for variable importance in these models supported H_3 assumptions, pointing at symptoms (classified with CCS method) as the most important variable -by far- in all the models, ranging from 69.13% to 85.64% for variable importance. Triage level followed consistently, staying in the interval 13.36%-25.97%, which was already a hint into H_4.

Triage method

The evaluation of the triage model effectiveness was performed with the admission rates for each triage level. The resulting plot, as seen in *Figure 8*, showed to be quite consistent with hypothesis H_4, supporting that the current triage model works reasonably well and is able to differentiate urgency level between patients. It only failed to show meaningful results for patients classified as level 4 and 5, the latter having a slightly higher rate of admissions. This can have several explanations, but given the numbers obtained from the complete dataset, showing five times more observations being classified as level 4 compared to level 5, it seems reasonable to assume that at the time of triage the professional behind the counter might be reluctant to assign a too low triage level to the patient if the symptoms are not very clear, and to play on the safe side assigns to such a visit a higher triage level than actually needed, which in the end might account for this mild inconsistency in the results.

Frequency of visits

To evaluate H_5, which hypothesized that a higher frequency of visits did not lead to higher admission rate, the correlation analysis already mentioned in *Results* was performed and no correlation was found with a very high significance level. This, however, did not help in the decision to accept or reject H_5. Thus, a table was built to show the number of accumulated visits next to the admission rate for each of the observed frequencies. From this table (see *Appendix I*), the frequency of admission graph was built. The results showed that the hypothesis was too short sighted in assuming that the behavior would be constant throughout the whole number of visits range. It is clear that the hypothesis is rejected for the range 1 to 5 visits/year, during which the admission rate increases from frequency to frequency, from 8.87% to 13.59%. After 5 visits/year, this rate starts to slowly decrease, but it does not go consistently below the rate observed for 1 visit/year until 26 visits/year, which showed an admission rate of 7.89% ad was followed by even lower rates. Before 26 visits/year, however, three frequencies (19, 20 and 24 visits/year) returned rates below 8.87%. Some higher frequencies returned surprisingly high admission rates (over 20% admission), but it would be statistically wrong to take them as significant given the low number of patients going to the ED for that large number of times (in no cases more than 2 patients during the 2018 exercise).

Something interesting to observe is the fact that for the top three highest frequencies (136, 147 and 244 visits) the admission rate was 0%. Diving into these three cases, all of them represented by only one different patient, we observe a common characteristic. The three patients show a clinical picture with anxiety at the center of it, with high rates of superficial injuries, connective tissue damage, and also very high residual/non-classified symptoms. This falls outside the scope of this research study, but it should serve as fundament for future research on how these patients suffering from anxiety can be better treated, and whether hospitals, and even more EDs, are the best access doors for patients with this type of disorders to the healthcare system.

Sensitivity analysis

This analysis targeted the development of prediction models with a sensitivity of at least 0.975.

As can be observed in *Table 4,* there is a tradeoff between sensitivity and specificity, with the latter quickly dropping when high levels of sensitivity are sought. As happened with the accuracy models, the approach that gives the best results is the one that uses the complete dataset to make predictions. With a specificity of 0.4262, it cannot be beaten by any of the other combinations. Even if the adults' dataset (18-115 years of age) shows a slightly better performance, when combined with the pediatrics counterpart they together give an averaged specificity of 0.4171, below that obtained with the complete sample. When trying to improve the adults' dataset performance with the two adult subsets, the same non-satisfactory result as with the accuracy models was found.

Nevertheless, the complete model shows better than expected results, and provides a good foundation for future research. Seeing that it is possible to predict 97.5% of the admissions while maintaining a specificity above 0.4, the researcher proposes to extend this piece of research to further improve crowding situations in EDs in Catalonia.

Based on the results (*Table 5*), it seems reasonable to assume that a lot of visits could be avoided with this approach if this information was available to the patient. From the 851,415 visits that ended up in discharge (from the test set in the complete dataset), 362,891 were accurately predicted. This is already a 37.93% of the total of visits in this test set.

	Reference	
Prediction	0	1
0	362891	2574
1	488524	102772

Table 5. Confusion matrix complete dataset

A model like that could allow for the development of a platform connecting patient and hospital prior to the former going to the ED. Said platform would ask patients for their identification information (either name and age, or national identification number, or hospital ID) and for the reasons behind their intention to go to the ED (symptoms or CCS). With this information it would be possible to run the patients' case through the predictive model and see how necessary it would be for each patient to go to the ED and how likely would it be that their situation shall require them to stay in the hospital. This, besides giving the hospital an idea of the number of patients that would be attending the ED in the next few hours, allowing the staff to prepare for the expected capacity and starting the bed allocation process with even more time in advance, would allow the patients to receive a very accurate prediction of the outcome of their visit to the ED. Moreover, this information could be combined with real-time information about the crowding situation in the ED which could be used by the patients to decide if they consider it worthy to go to the ED and wait X hours there until being attended, or otherwise they prefer to wait for a few hours or even until the next day to go to the hospital -as long as their situation does not need immediate attention- in order to lose less time waiting to be attended, in case the current situation in the ED is of high crowding and long waiting times are expected.

Future research should focus on identifying how many of these patients correctly identified as discharges, in situations of high crowding in the ED, and after being provided with the information about the >97.5 probability of their visit ending in discharge (complete dataset probability of admission threshold = 0.022) and the information about long waiting time at the ED before being visited by the doctor, would decide to postpone their visit and therefore contribute to reduce overcrowding.

Implications

Previous studies have explored possibilities in the admission prediction field, but many of them have done so with data from trauma level I centers exclusively, and the ones overcoming the specialization frontier have only provided results at an individual hospital level. This is the first study to do predictive analysis for a comprehensive hospital network including EDs of diverse size and characteristics, and it has achieved better performing results in the conditions studied.

It is envisioned, given the positive results obtained with the complete dataset for both the AUC and sensitivity analyses, that the Catalan Healthcare System will acknowledge the potential of integrating these models into the daily operations of EDs in the Integral Healthcare System for Public Use in Catalonia (SISCAT) to provide professionals with a powerful tool to forecast upcoming admissions in advance and start the bed allocation process at the same time as the visit takes place, overcoming the main inefficiency of the bed allocation process, which is the unnecessary presence of ready-to-be-admitted patients in the ED that actively contribute to ED overcrowding. Moreover, this would reduce the negative impact of ED overcrowding in terms of long patient waiting times, decreased patient satisfaction, or inadequately treated pain, among others. The benefits for patients and professionals are clear, but probably more relevant to the governmental entities would be the economic savings resulting from an optimized ED service, which previous studies have found to be around 350€ for every hour a patient stays in the onboarding process into the hospital [36,37]. This figure scaled to a whole regional hospital network managing more than 350,000 admissions in a yearly basis could translate into enormous savings in the millions range. A proper financial analysis would be required to get a better and precise estimation of the actual figures.

Limitations

The results obtained should be interpreted considering five main limitations. The first of them is the fact that the models obtained performed significantly well, but it has not been determined whether this was a result of the quality of the dataset, or the GBM method used. Future research should focus on the application of existing predictive techniques -such as logistic regression- on this same dataset to better determine how beneficial is it to use a dataset including data from a large healthcare system against one including data from an individual center, and how better are the results obtained with GBM when compared to those obtained with other predictive techniques.

The second limitation, which has already been mentioned in accuracy testing results, is the fact that the models used in the study did not include patient history information, even though previous literature proved its beneficial contribution in the performance of predictive models for hospital admissions. This was an unexpected setback for the study given that the researcher asked for data related to number, symptoms, lab tests, and admission results of previous visits for the patients that attended the ED during the 2018 exercise, but the Covid-19 situation made it impossible for the Catalan Healthcare System to treat, organize and provide this extra information, alleging a too large amount of information to be managed and a very complex task of elaboration of the dataset for analysis, which would require more resources than available during the coronavirus crisis. It would be therefore interesting to include patient history information in future studies with the goal of improving the AUC for the models to be used in EDs at time of arrival, and to increase the performance of the sensitivity-oriented models.

Another limitation regarding the data used for the model is the lack of information about date and time of arrival for the visits received at the ED. This was a piece of information that was also formally requested, but that could not be provided due to the hazard to properly and accurately integrate it with the rest of the variables. Therefore, it was not possible in this research to establish a chronological order between the visits observed for each of the patients, which could have otherwise allowed for the creation of an extra variable reporting the position of each visit inside the total number of visits for the patient during the year and could have potentially improved the results. The effects of this missing information were attempted to be minimized with the creation of the variable 'accumulated number of visits', but as the variable importance results showed (*Appendix IV*), it did not help improve much the performance of the models.

Additionally, the models developed in this study did not account for the obvious higher admission rates that some symptoms show when compared to other, often not-as-serious, symptoms. As it has been seen, the symptom with the highest admission rate was polyhydramnios and other disorders in the amniotic cavity, which makes a lot of sense since it is an indicator for a woman that is going to be giving birth imminently. These kinds of symptoms might bias the real performance of the model to the positive side, since overcrowding comes, in many cases, from the uncertainty of a patient being admitted in the hospital or not, rather than from visits with a >90% likelihood of ending in admission. To solve this limitation in future research, it is proposed to develop models including only those patients showing symptoms that have a rate of admission below a certain percentage (it has not been possible to find any literature that would help determine what would be an adequate threshold), and train and test these models to see how good they could perform in this more complex situation.

Finally, this study gives the first foundation to sensitivity-aimed models and proposes the integration of such models in a platform that would allow patients to remotely get a first impression of how important it would be for them to go to the ED, and would also make it possible for them to evaluate personally whether their visit could be postponed -in cases of ED crowding and long waiting times. However, no further research has been done neither in the patient or the hospital side. To draw significant conclusions about the possible impact and implementation of a platform like the one described it would be necessary to first get an idea of how costly it would be for hospitals to get and share real-time information about the situation at their EDs, how difficult would the integration of this data with the sensitivity-predictive-models' software be, and how many visits could potentially be avoided with this approach. Only after having conducted this analysis thoroughly would it be possible to give an accurate prediction of the potential impact of this method towards solving ED overcrowding.

Economic Analysis

Given the nature and conditions of the study, the economic requirements behind it have been very few. There have not been any product development nor major software expenses, so the only costs incurred have been these related to the professional hours of work invested both by the researcher and the academic supervisor, which add to a total of 9,200€.

The detail of all these costs is displayed in *Table 6* included in the *Budget.*

Conclusions

This study reaffirms on the belief that gradient boosting machine is a powerful tool to use in binary outcome predictive models. It shows that data collected at the moment of arrival to the ED can be used to predict hospital admissions accurately, and that a model including data from a comprehensive hospital network has a better predictive performance when compared to a similar model developed with data from one unique health center only. It discusses the huge potential that the application of the models obtained could have in fighting crowding in EDs by allowing for an early start of the bed allocation process, making it possible to do all the required procedures for admission simultaneously to the patient being visited by the doctor, instead of doing it in a sequential manner after the visit, which unnecessarily crowds ED rooms and generates a non-optimal use of the available resources in EDs. The study also suggests the application of this predictive technique to develop models with proven high sensitivity to digitalize the patient-hospital relationship in order to allow for a first contact between both parties before the visit to the ED, which can potentially regulate the inflow of patients in this department and reduce ED overcrowding significantly.

References

1. *Hospital-Based Emergency Care: At the Breaking Point.* Institute of Medicine 2007. Washington, DC: The National Academies Press.

2. *Unidad de Urgencias Hospitalarias. Estándares y Recomendaciones.* Ministerio de Sanidad y Política Social. Informes, Estudios e Investigación 2010. Madrid.

3. Franaszek, J. B., Asplin, B. R., Brunner, B., Epstein, S. K., Fields, W. W., Hill, M. B., ... & Schneider, S. M. (2002). Responding to emergency department crowding: A guidebook for chapters. *Technical report, American College of Emergency Physicians, Irving, TX*.

4. Jayaprakash, N., O'Sullivan, R., Bey, T., Ahmed, S. S., & Lotfipour, S. (2009). Crowding and delivery of healthcare in emergency departments: the European perspective. *Western Journal of Emergency Medicine*, *10*(4), 233.

5. Sun, B. C., Hsia, R. Y., Weiss, R. E., Zingmond, D., Liang, L. J., Han, W., ... & Asch, S. M. (2013). Effect of emergency department crowding on outcomes of admitted patients. *Annals of emergency medicine*, *61*(6), 605-611.

6. Bernstein, S. L., Aronsky, D., Duseja, R., Epstein, S., Handel, D., Hwang, U., ... & Schafermeyer, R. (2009). The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*, *16*(1), 1-10.

7. JJ, A. (2016). Long emergency department boarding times drive walkaways, revenue losses. ACEP Now.

8. Medley, D. B., Morris, J. E., Stone, C. K., Song, J., Delmas, T., & Thakrar, K. (2012). An association between occupancy rates in the emergency department and rates of violence toward staff. *The Journal of emergency medicine*, *43*(4), 736-744.

9. Barak-Corren, Y., Fine, A. M., & Reis, B. Y. (2017). Early prediction model of patient hospitalization from the pediatric emergency department. Pediatrics, 139(5), e20162785.

10. Amarasingham, R., Swanson, T. S., Treichler, D. B., Amarasingham, S. N., & Reed, W. G. (2010). A rapid admission protocol to reduce emergency department boarding times. BMJ Quality & Safety, 19(3), 200-204.

11. Howell, E. E., Bessman, E. S., & Rubin, H. R. (2004). Hospitalists and an innovative emergency department admission process. Journal of general internal medicine, 19(3), 266-268.

12. Romero, F. B., Macías, J. B., Gil, D. G., & Álvaro, J. L. (2010). Tiempo de demora para la hospitalización tras la implantación del ingreso directo a cargo del Servicio de Urgencias. Revista clinica espanola, 210(4), 159-162.

13. Alonso, D. G., Enguix, N., Valverde, L., Castells, M., Pascual, I., Esquerda, A., & Sarlé, J. (2011). Resultado de un proceso para la mejora de las altas hospitalarias precoces. Emergencias: Revista de la Sociedad Española de Medicina de Urgencias y Emergencias, 23(1), 29-34.

14. Ovens, H. (2010). Saturación de los servicios de urgencias: Una propuesta desde el sistema para un problema del sistema. Emergencias: Revista de la Sociedad Española de Medicina de Urgencias y Emergencias, 22(4), 244-246.

15. Worster, A., Gilboy, N., Fernandes, C. M., Eitel, D., Eva, K., Geisler, R., & Tanabe, P. (2004). Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. Canadian Journal of Emergency Medicine, 6(4), 240-245.

16. Ebrahimi, M., Heydari, A., Mazlom, R., & Mirhaghi, A. (2015). The reliability of the Australasian Triage Scale: a meta-analysis. World journal of emergency medicine, 6(2), 94.

17. Beveridge, R., Ducharme, J., Janes, L., Beaulieu, S., & Walter, S. (1999). Reliability of the Canadian emergency department triage and acuity scale: interrater agreement. Annals of emergency medicine, 34(2), 155-159.

18. Mackway-Jones, K., Marsden, J., & Windle, J. (Eds.). (2014). Emergency triage: Manchester triage group. John Wiley & Sons.

19. Soler, W., Gómez Muñoz, M., Bragulat, E., & Álvarez, A.. (2010). El triaje: herramienta fundamental en urgencias y emergencias. Anales del Sistema Sanitario de Navarra, 33(Supl. 1), 55-68.

20. Penades, M., Ros, I. (2015). Forecasting patients' admissions in an ED: The case of the Meyer Hospital. Università degli studi Firenze.

21. Leegon, J., Jones, I., Lanaghan, K., & Aronsky, D. (2005). Predicting hospital admission for Emergency Department patients using a Bayesian network. In AMIA... Annual Symposium proceedings. AMIA Symposium (Vol. 2005, pp. 1022-1022). American Medical Informatics Association.

22. Leegon, J., Jones, I., Lanaghan, K., & Aronsky, D. (2006). Predicting hospital admission in a pediatric Emergency Department using an Artificial Neural Network. In AMIA Annual Symposium Proceedings (Vol. 2006, pp. 1004-1004). American Medical Informatics Association.

23. Savage, D. W., Weaver, B., & Wood, D. (2017). P112: Predicting patient admission from the emergency department using triage administrative data. Canadian Journal of Emergency Medicine, 19(S1), S116-S116.

24. Lucke, J. A., de Gelder, J., Clarijs, F., Heringhaus, C., de Craen, A. J., Fogteloo, A. J., ... & Mooijaart, S. P. (2018). Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. Emerg Med J, 35(1), 18-27.

25. Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., ... & Kelen, G. (2018). Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Annals of emergency medicine, 71(5), 565-574.

26. Hong WS, Haimovich AD, Taylor RA (2018) Predicting hospital admission at emergency department triage using machine learning. PLoSONE 13(7): e0201016

27. Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

28. Mayr, A., & Hofner, B. (2018). Boosting for statistical modelling-A non-technical introduction. Statistical Modelling, 18(3-4), 365-384.

29. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

30. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

31. Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. The R Journal, 9(1), 421-436.

32. *Gradient Boosting Machines.* (2018). UC Business Analytics. R Programming Guide. University of Cincinnati.

33. *Decret 196/2010, de 14 de desembre, del sistema sanitari integral d'utilització pública de Catalunya (SISCAT).* Portal Jurídic de Catalunya. Servei Català de la Salut (CatSalut). Generalitat de Catalunya.

34. *Clinical Classification Software (CCS) 2015.* Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP).

35. DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics, 44(3), 837-845.

36. Falvo, T., Grove, L., Stachura, R., Vega, D., Stike, R., Schlenker, M., & Zirkin, W. (2007). The opportunity loss of boarding admitted patients in the emergency department. Academic Emergency Medicine, 14(4), 332-337.

37. Pines, J. M., Batt, R. J., Hilton, J. A., & Terwiesch, C. (2011). The financial consequences of lost demand and reducing boarding in hospital emergency departments. Annals of emergency medicine, 58(4), 331-340.