scientific reports

OPEN

Investigation of emergency department abandonment rates using machine learning algorithms in a single centre study

Marta Rosaria Marino[1], Teresa Angela Trunfio[2]✉, Alfonso Maria Ponsiglione[3], Francesco Amato[3] & Giovanni Improta[1,4]

A critical problem that Emergency Departments (EDs) must address is overcrowding, as it causes extended waiting times and increased patient dissatisfaction, both of which are immediately linked to a greater number of patients who leave the ED early, without any evaluation by a healthcare provider (Leave Without Being Seen, LWBS). This has an impact on the hospital in terms of missing income from lost opportunities to offer treatment and, in general, of negative outcomes from the ED process. Consequently, healthcare managers must be able to forecast and control patients who leave the ED without being evaluated in advance. This study is a retrospective analysis of patients registered at the ED of the "San Giovanni di Dio e Ruggi d'Aragona" University Hospital of Salerno (Italy) during the years 2014–2021. The goal was firstly to analyze factors that lead to patients abandoning the ED without being examined, taking into account the features related to patient characteristics such as age, gender, arrival mode, triage color, day of week of arrival, time of arrival, waiting time for take-over and year. These factors were used as process measures to perform a correlation analysis with the LWBS status. Then, Machine Learning (ML) techniques are exploited to develop and compare several LWBS prediction algorithms, with the purpose of providing a useful support model for the administration and management of EDs in the healthcare institutions. During the examined period, 688,870 patients were registered and 39188 (5.68%) left without being seen. Of the total LWBS patients, 59.6% were male and 40.4% were female. Moreover, from the statistical analysis emerged that the parameter that most influence the abandonment rate is the waiting time for take-over. The final ML classification model achieved an Area Under the Curve (AUC) of 0.97, indicating high performance in estimating LWBS for the years considered in this study. Various patient and ED process characteristics are related to patients who LWBS. The possibility of predicting LWBS rates in advance could be a valid tool quickly identifying and addressing "bottlenecks" in the hospital organization, thereby improving efficiency.

Keywords Emergency department, Overcrowding, Leave without being seen, Machine learning

Overcrowding in emergency rooms has been acknowledged as a critical challenge in hospital administration[1]. The rising demand for Emergency Departments (EDs) services has become a global healthcare issue, obtaining more attention in recent years. Overcrowding in the ED is characterized by a scenario in which the demand for services exceeds the capacity of healthcare providers (nurses, physicians, etc.), leading to the inability of offer appropriate care in a timely manner. The prolonged stays of admitted patients in the ED for extended durations is one of the primary causes of ED congestion. A high volume of patients in the ED, that outweighs its size, results in delays in diagnosis and treatment, as well as in the inefficiency of care services and the reduction of patient satisfaction[2]. One of the main objectives of first aid operations is to minimize the patient's initial wait, defined as the period between registration and the start of the evaluation by a doctor. The Emergency Severity Index (ESI)

[1]Department of Public Health, University of Naples "Federico II", Naples, Italy. [2]Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy. [3]Department of Electrical Engineering and Information Technologies, University of Naples "Federico II", Naples, Italy. [4]Interdepartmental Center for Research in Healthcare Management and Innovation in Healthcare (CIRMIS), University of Naples "Federico II", Naples, Italy. ✉email: teresa.trunfio@gmail.com

of the triage system, which classifies patients according to a priority scale from 1 to 5, suggest that patients on treatment must be taken care of within 1, 10, 30, 60 and 120 min respectively[3]

It is necessary to prevent a linear increase in ED length of stay (LOS)[4]. The variables that contribute to longer LOS for patients admitted in the ED have been thoroughly documented, and among the most challenging to address.

Starting from this consideration, it is easy to understand that a large number of patients have reported waiting in the ED for extended LOS before deciding to leave without being seen (LWBS)[5]. Patients LWBS are those who are registered and triaged for care but then they leave without any visit by a physician[6]; it is recognized as a key indicator for monitoring EDs overcrowding and occupancy, which are difficult to evaluate directly. This patient population has been well-studied as a useful marker of overcrowding. Studies have shown that nearly half of LWBS patients may need immediate or urgent medical evaluation and treatment, and 1 in 20 of these patients requires hospitalization. Although most of the patients who LWBS do not have urgent medical problems, they still require medical consideration. This condition compromises both patient experience and ED safety, risk management and patient care efficiency[7].

Quality of care in healthcare contexts, as also demonstrated in various studies[8–11], is associated with variations in key indicators related to the patient status and the healthcare process, including LWBS. In fact, high rates of patients LWBS negatively affect healthcare institutions' profits resulting in a significant financial loss[12]. An 'acceptable' LWBS rate is approximately 2 to 3%[13].

Identifying conditions strongly associated with high rates of LWBS could help minimize the occurrence of this problem. To this regard, a predictive model would be useful to identify such factors and planning for subsequent improvements in resources allocation and manpower, aiming to reduce the number of LWBS patients.

This work continues and extends a previously published conference study in which a limited number of years and variables were used to investigate this phenomenon[14].

Literature review

In both literature and clinical practice, patients LWBS have been identified and studied as a useful predictor for the ED efficacy and effectiveness because they facilitate the evaluation of the patient satisfaction and treatment quality[15].

Numerous researchers have examined how to estimate this critical value using innovative and sophisticated analytic approaches and artificial intelligence techniques[16–21]. Several studies have highlighted the potential of identifying new clinical pathways for various healthcare operations practices through data processing and Machine Learning (ML) techniques. These approaches have been recognized as powerful solutions for enhancing clinical management standards and quality of care[21–31]. Researchers have investigated the relationship between patients who LWBS and their features including age, gender, and comorbidity, as well as ED process parameters like time of registration and assigned triage color[32–37]. Further studies investigated the ED abandonment rates through computational methods and mathematical models, some of which are based on logistics notions drawn from the corporate world[38–43].

The purpose of this study is to examine the reasons contributing to the rise in LWBS patients in the ED, and to develop a classification algorithm for distinguishing LWBS patients from those who will be assessed by a physician. Various patient-related data, such as age, gender, LOS, and triage score, were used as predictor factors. The research was conducted collecting data from patients enrolled at the ED of the University Hospital of Salerno "San Giovanni di Dio e Ruggi d'Aragona" (Italy). Previous studies in the literature typically use national databases and are limited to the study of factors, using logistic regression models or simple statistical analyses. In contrast to other studies, the availability of a large amount of data from a single hospital allowed us to appropriately train several sophisticated ML algorithms[44–46]. Although a single-centre study limits the generalizability of the results, it enables the translation of findings into specific corrective actions that can be directly adopted by the hospital. The predictive algorithms and the Firth logistic regression—not a simple multivariate one but one that specifically targets unbalanced datasets—were implemented following a correlation analysis of the input parameters. The evaluation metrics of the four ML algorithms considered were tested and compared to obtain a high-performance prediction model.

Aim of the study

In this study, we use a predictive model based on the hypothesis that there are multiple independent hospital factors related to the increase in LWBS rate in the ED. Initially, we aimed to identify which patient and ED process-related aspects have the most significant impact on LWBS. The capture of the features of a large number of patients provided the opportunity of a consistent data analysis. The knowledge gained from this phase enable the implementation of corrective measures to mitigate this phenomenon, which, as discussed, can lead to serious consequences in patients who no longer receive proper care. Then, we built a classification model to predict LWBS using ML algorithms. Through these models, it will be possible to know in advance whether there are boundary conditions that could generate dissatisfaction and thus abandonment of the ED. In addition, through a study of feature importance it will be possible to enter into the prediction and further highlight the relationships between the factors.

Methods

Data collection and dataset features

This retrospective study was carried out analysing the records from the Emergency Department database from the "San Giovanni di Dio e Ruggi d'Aragona" University Hospital, Salerno (Italy). Data for ED patients registered from 2014 to 2021 were extracted from the ED database, to offer more consistent estimates. The daily data of all patients who logged into the ED, those who were hospitalized, and those who LWBS during the study period were analysed.

Thus, we identified our cohorts using a binary value for the presence or absence of LWBS status. This value was generated for all ED patients since they were all triaged, but some left the ED without being evaluated by a healthcare provider. Starting with the examination of 688,870 records related to ED registrations, the dataset was properly processed to ensure compatibility with the ML algorithms implemented.

The following information were considered for each patient:

- Gender,
- Age;
- Access mode, divided into 3 classes:
    - autonomous, that includes patients arriving to the ED by themselves;
    - via ambulance, that includes patients arriving to the ED by 118 ambulance or other ambulances like Army's ones, fire brigade's ones or from other regions;
    - undisclosed;
- Triage score, divided by colors according to the patient's condition:
    - red;
    - yellow;
    - green;
    - white;
- Time of arrival, divided into 5 classes:
- 00:00 – 05:59;
- 06:00 – 11:59;
- 12:00 – 17:00;
- 18:00 – 20:59;
- 21:00 – 23:59;
- Day of week of arrival;
- Waiting time for take-over in minutes, for patients who leave before take-over the total time in the ED will be used;
- Mode of discharge:
- To home;
- To outpatient facilities;
- Abandonment before closure of ED records;
- LWBS;
- Admission to the ward;
- Refuses admission;
- Transfer to facilities in the territory;
- Transfer to another institution.

The dataset is detailed in Table 1.

Patients arriving in the ED already deceased (triage color Black) and those who died in the ED were excluded from the analysis. In addition, patients for whom age, gender and triage color was not reported upon admission, as well as those discharged with ED records cancelled or closed by the case manager months later, were eliminated.

The distribution of accesses by color code and year is shown in Fig. 1.

The highest number of accesses was recorded in 2020 (103,537), distributed across all complexity classes.

From an operational point of view, the variables Gender, Access Mode, Time of Arrival and Day of week of arrival were coded as dummy variables, i.e. they were decomposed into n dummy variables made of 0/1 where n is the number of possible alternatives defined for each identified feature. The dependent variable was extracted from the Mode of Discharge feature, setting 1 for LWBS patients and 0 otherwise. To explore the differences between patients who LWBS and those who did not, a correlation analysis was performed between patients' characteristic. Using Pearson's correlation, we studied any correlations present among the independent variables before proceeding with the analysis. Then, through these features related to patients, a model was built by means of a ML approach to predict LWBS status. The models were applied to data related to each year, in order to evaluate its performance with variations in the features of the ED patients.

Data analysis

Once the records of interest were collected from the company information system, the data were analyzed. The aim was to study the drop-out phenomenon through the influence of certain variables associated with the patient and their flow in the ED.

First, the dataset was divided into two groups:

- LWBS patients;

Variable	N = 688,870
Gender	
Male	361,054
Female	327,816
Age	
Mean	43.89
Median	44.00
Standard Deviation	25.03
Access mode	
Autonomous	548,190
Via Ambulance	121,823
None	18,857
Triage score	
Red	8497
Yellow	94,464
Green	557,764
White	28,145
Time of arrival	
00:00–05:59	61,309
06:00–11:59	258,106
12:00–17:59	221,885
18:00–20:59	103,389
21:00–23:59	44,181
Day of week of arrival	
Monday	110,028
Tuesday	98,560
Wednesday	97,568
Thursday	97,666
Friday	98,344
Saturday	97,344
Sunday	89,360
Waiting time for take-over (min)	
Mean	51.66
Median	28.00
Standard deviation	90.89
Mode of discharge	
To home	271,989
To outpatient facilities	208,706
Abandonment before closure of ED records	23,347
LWBS	39,188
Admission to the ward	107,940
Refuses admission	32,579
Transfer to facilities in the territory	35
Transfer to another institution	5086

Table 1. Dataset features.

- no-LWBS patients.

These two groups were first compared through statistical analysis. For continuous variables, such as age and waiting time, the U-Mann Withney test was used, while for dichotomous variables, such as gender, mode of access and others, the Chi-squared test was applied. For both, a p-value < 0.05 showed the significance of the comparison.

In addition to using Logistic Regression (LR) as a classifier, which will be discussed in the next section, Firth-type penalization was implemented. This method is widely used to reduce the small-sample bias of maximum likelihood coefficients. On unbalanced datasets like the one under consideration, maximum likelihood estimation could lead to either no results or highly biased results. This method allows for more accurate estimates with minimal bias[47].

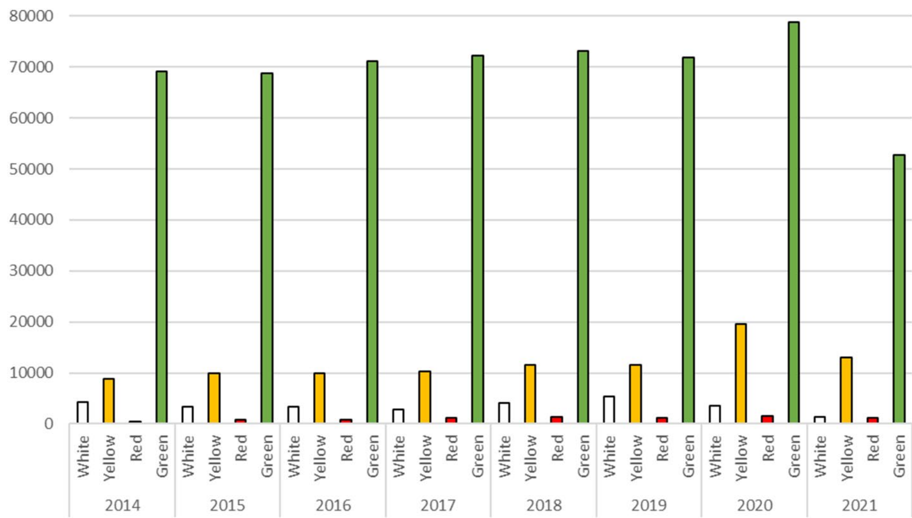

Fig.1. Distribution by year and triage color.

The Google Colab cloud computing environment and the Python programming language were used for the analyses, with the exception of the Firth logistic regression implemented in R Studio.

Classification algorithms

Machine Learning (ML) algorithms are learning function that map input variables to an output value to make predictions. This general learning task allows making future predictions given new samples of the same input variables, thus creating a system capable of understanding and improving performances based on the data analysis.

In this study, supervised classifiers were implemented. Several independent variables were used as input variables for the ML algorithms. The LWBS status was converted into a string variable and used as the output, with two clusters labelled for classification (leave and not leave). To perform the classification analysis for predicting and discriminating patients who will LWBS, the capabilities of different classification algorithms were exploited. Specifically, the four selected algorithms are: Random Forest (RF), Naïve Bayes (NB), Decision Tree (DT) and Logistic Regression (LR). All algorithms were implemented using the scikit-learn library[48].

RF is a type of supervised ML algorithm based on learning multiple predictive models to form a single, more powerful prediction model. Each model used by the RF prediction is usually a decision tree (DT). NB classifier is a probabilistic classification algorithm based on Bayes' Theorem. It requires a strong assumption of independence between features but has demonstrated good performance in binary classifications.

LR is a probabilistic classification algorithm based on supervised learning that builds a data classification model through a sigmoid function that converts the predictions to probabilities. Specifically, LR evaluates the relationship between the input variables and the output, estimating the probabilities that inputs belong to a specific class through the logistic function to assign the result to a class based on the probability. Commonly, it is used for binary classification.

For all algorithms, the dataset was divided in a training set (80%) and a test set (20%). The performance of the algorithms was evaluated according to the following parameters:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{1}$$

$$\Pr ecision = \frac{TP}{TP + FP}. \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3}$$

$$F1Score = 2 * \frac{\Pr ecision * \text{Recall}}{\Pr ecision + \text{Recall}}. \tag{4}$$

$$Balanced\ Accuracy = \frac{\frac{TP}{TP+FP} + \frac{TN}{TN+FN}}{2}, \tag{5}$$

where:

- TP: True Positive, a person LWBS and classified as LWBS (class 1);

- TN: True Negative, a person no-LWBS and classified as no-LWBS (class 0);
- FP: False Positive, a person no-LWBS and classified as LWBS (class 0–1);
- FN: False Negative, a person LWBS and classified as no-LWBS (class 1–0).

In addition to these parameters, due to the dataset being unbalanced, the AUC 'Area under the Receiver Operating Characteristic (ROC) curve' was also used. The ROC curve shows the performance of the classification algorithms, graphing Recall and the false positive rate defined as the ratio of FP to FP + TN. The AUC measures the entire two-dimensional area under the ROC curve (integral calculation) from (0, 0) to (1, 1). To obtain results that are not dependent on a particular partition performed, Cross Validation[49] with cv = 10 was implemented, combined with GridSearchCV tool[50] to search for optimal hyperparameters. GridSearchCV tests all possible combinations of a range of values given for each hyper-parameter (Table 2). For each combination, it evaluates the model using the Cross-Validation method. This process yields performance results for each hyperparameter combination, allowing us to choose the one with the best performance. Given the dataset's high imbalance, AUC was selected as the observation performance variable to guide the optimization process.

Prior to partitioning, the Synthetic Minority Oversampling Technique (SMOTE)[51] was applied in order to duplicate the examples in the minority class (LWBS patients are in any case a low % of the total). Though these examples do not add new information to the model, they help counteract the imbalance. Additionally, DT, RF and LR algorithms were selected because they were known in the literature to perform well in these cases[52–54] and it was possible to set class_weight: 'balanced' to automatically adjust weights inversely proportional to their frequency in the input dataset. NB, on the other hand, was used as a reference for comparing performance. A further statistical test was assessed the stability of the model by checking for statistically significant variation in accuracy when randomly permuting some samples of the test set 100 times[55]. To determine the best algorithm, McNemar's test was implemented to compare the confusion matrix of the algorithm with the highest AUC with others, using a significance level of 0.05. Four values[56] was calculated on which the chi-squared test was implemented:

- The number of instances misclassified by both classifiers (C_{00});
- The number of instances misclassified by the first classifier but correctly classified by the second (C_{01});
- The number of instances misclassified by the second classifier but correctly classified by the first (C_{10});
- The number of instances correctly classified by both classifiers (C_{11}).

After identifying the best algorithm, the classification process was analyzed using Feature Importance. Again, an iterative procedure called Permutation Feature Importance was used, which replaces one independent variables with a corrupted version at each iteration. The loss of performance for each iteration was assessed by measuring the AUC. The result, shown graphically, indicates the loss in AUC for each variable, revealing the importance of each feature.

Ethics approval and consent to participate

The authors declare that all methods were performed in accordance with the Declaration of Helsinki.

The institutional review board of "San Giovanni di Dio and Ruggi d'Aragona" University Hospital has approved the study.

The institutional review board of "San Giovanni di Dio and Ruggi d'Aragona" University Hospital provided waiver for informed consent for the study.

Our data, provided by the Hospital's Health Department, are completely anonymous and no personal information are linked or linkable to a specific person.

Results

Over the 8 years studied, 688,870 patients were registered and 39,188 (5.68%) left without seeing a physician. The detail by year is shown in Fig. 2.

Figure shows that the lowest number of LWBS patients occurred in 2014 (4.70%), while the highest was in 2019 (6.77%). A first way to study this phenomenon is through statistical analysis. Two groups were created based on discharge mode (LWBS No / LWBS Yes) and compared using the U-Mann Whitney test and the Chi-squared test according to the type of variable. The results are presented in Table 3.

For all variables, there was a statistically significant difference between the two distributions. Males, with a higher median age, who access on Mondays, and those with a White triage code were more likely to leave without being seen. Then, a correlation study was conducted and the results are shown in Fig. 3.

Algorithms	Hyperparameters
NB	'var_smoothing': np.logspace(0,-9, num = 100)
RF	'n_estimators': [5, 10, 15, 20], 'max_depth': [2, 5, 7, 9]
DT	'max_depth':range(3,20)
LR	'C':np.logspace(-3,3,7), 'penalty':['l1', 'l2']

Table 2. Selected values of each hyperparameter.

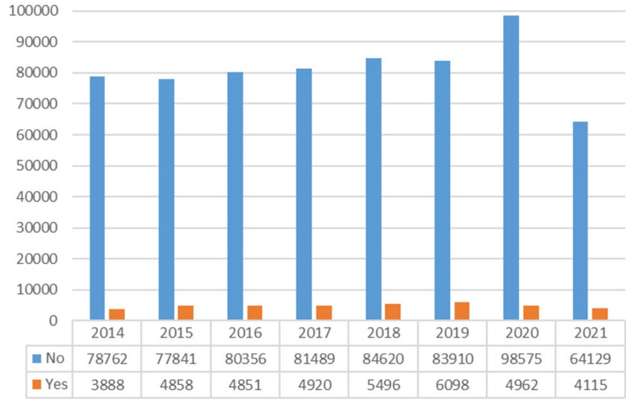

Fig.2. Number of LWBS patients in total and per years.

The highest correlations were found between Age and Triage Score, Access Mode-Autonomous and Triage Score, Gender Male and Female, and various access modes and times. This last result is particularly interesting as it demonstrates that LWBS patients wait significantly longer before being attended to by ED physicians. Box plots in Fig. 4 reports that LWBS patients experience longer wait times and greater variability, with numerous outliers in both distributions, detailed in Table 4.

In the case of LWBS patients, the box had a higher mean and a larger amplitude in the presence of greater variability, taking into account the large number of outliers present in both distributions. Both distributions had a common maximum wait time, but an increasing value in minutes was found for patients who eventually left the ED.

Finally, machine learning (ML) algorithms were evaluated to predict LWBS occurrences and determine which algorithm performed best. The metrics used, with an emphasis on the AUC value for comparing classifiers, are shown in Tables 5, 6.

The results indicate that, for the considered dataset, the selected classifiers have a similar AUC, but the classification algorithm that achieves the highest accuracy and AUC value was the Decision Tree. Indeed, this classifier reaches and accuracy of 94.00%. Therefore, McNemar's test was implemented to check whether there is a significant difference between the performance of DT and the other algorithms by studying the errors and correct classifications in the confusion matrix. For all of them, the p-value was less than 0.05, demonstrating DT's better performance. In addition, the value of the statistical test below the threshold of 0.05 reported in Table 6 shows that the score varies significantly as the test set changes, demonstrating the stability of the model.

The ROC curve for the best algorithm is shown in Fig. 5.

To assess the impact of single variables on the model, Feature Importance was studied, as shown in Fig. 6.

Gender-Female, Gender-Male, Access Mode-Autonomous, Time Arrival-06:00–11:59 and 12:00–17:00 were the variables that most affected the prediction. Starting from the model prediction, improvement actions could be implemented in order to reduce the LWBS rates. First, staffing could be adjusted based on the volume of patients in the ED to improve both the triage phase and the taking care procedure. Moreover, the model results could be analysed to develop new protocols for enhancing triage workflow and standardizing the decision process for beds use and placement of patients depending on their criticality.

Finally, the Logistic Regression of Firth was implemented. Table 7 shows the obtained results.

As obtained for the statistical analysis, Gender, Waiting Time, Time Arrival-00:00–05:59 and 18:00–20:59, Age, Triage and Access Mode significantly influenced the decision to leave the ED.

Discussion

The Emergency Department (ED) serves as the interface between the hospital and critically ill patients. In recent years, ED usage has steadily increased, often surpassing the department's capacity to provide timely, quality care[7,57]. Indeed, in the context of ED management, the monitoring of the quality of services is crucial. Therefore, a number of performance indicators have been proposed to investigate and control the efficiency and effectiveness of ED workflows. One key indicator is the LWBS rate, which reflects patients who register and triage for care but leave without being seen by a physician. This issue is prevalent in many EDs due to the increased demand for healthcare services and overcrowding. LWBS has become an efficient indicator for timeliness and efficiency of ED services, as it highlights the ED's ability to manage patient intake and reduce wait times. Notably, 70% LWBS patients return to EDs in the following 24 h, and 11% LWBS require hospitalization within 7 days of their initial visit.

Long waiting times, overcrowding, lack of shared and standardized communication and information protocols and procedures, as well as patients' dissatisfaction towards healthcare services are leading cause of LWBS phenomenon. Therefore, a proper study of LWBS rates could represents an opportunity to improve several aspects and factors causing a decrease in the performance of ED processes.

Unlike other studies, this paper leverages a substantial dataset to develop a significant prediction model for the LWBS phenomenon. More precisely, 8 patient-related variables, i.e., age, gender, arrival mode, triage color,

Variable	LWBS No N = 649,682	LWBS Yes N = 39,188	p-value
Gender			**0.000**
Male	337,422	23,632	
Female	312,260	15,556	
Age			**0.000**
Median	43.00	45.00	
Standard deviation	25.22	21.54	
Access mode			**0.000**
Via ambulance	117,371	4452	
Autonomous	514,605	33,585	
None	17,706	1151	
Triage Score			**0.000**
Red	8464	33	
Yellow	93,456	1008	
Green	523,888	33,876	
White	23,874	4271	
Time of arrival			**0.000**
00:00–05:59	58,349	2960	
06:00–11:59	243,315	14,791	
12:00–17:59	208,929	12,956	
18:00–20:59	97,521	5868	
21:00–23:59	41,568	2613	
Day of week of arrival			**0.000**
Monday	102,374	7654	
Tuesday	92,632	5928	
Wednesday	91,989	5579	
Thursday	92,110	5556	
Friday	92,876	5468	
Saturday	92,321	5023	
Sunday	85,380	3980	
Waiting time for take-over (min)			**0.000**
Median	27.00	80.00	
Standard deviation	74.94	209.36	
Year			**0.000**
2014	78,762	3888	
2015	77,841	4858	
2016	80,356	4851	
2017	81,489	4920	
2018	84,620	5496	
2019	83,910	6098	
2020	98,575	4962	
2021	64,129	4115	

Table 3. Statistical comparison of the two groups. Significant values are in bold.

day of week of arrival, time of arrival, waiting time for take-over and year, were used to predict LWBS by means of ML algorithms.

Previous studies by Rathlev et al.[16], Pham et al.[17] and Tropea et al.[18] indicate that patients who wait the longest, have less severe codes and are younger are more likely to drop out the ED. Our study partially confirms these findings, except for age, where LWBS patients had a higher median age. Consistent with Sheraton et al.[21], male patients and those accessing on weekdays are more likely to leave. All of the above studies further extend the range of variables considered, including variables such as ethnicity or the type of clinical need that led to access, which cannot be compared with our study, as they are not present in our analysis. Nevertheless, despite the fact that the studies have been carried out in different countries and in different health systems, some features of the phenomenon remain common.

Our model's strengths include the large sample size used for training and the resultant predictive power. Indeed, even if the dataset was cleaned deleting all the records extracted from the database with missing variables, we collected a total of 688,870 records, catching 39,188 (5.68%) patients who LWBS. As shown, the performance reaches by DT classifier tool is remarkable, with an accuracy of 94.00%. The fact that a tree classifier performs

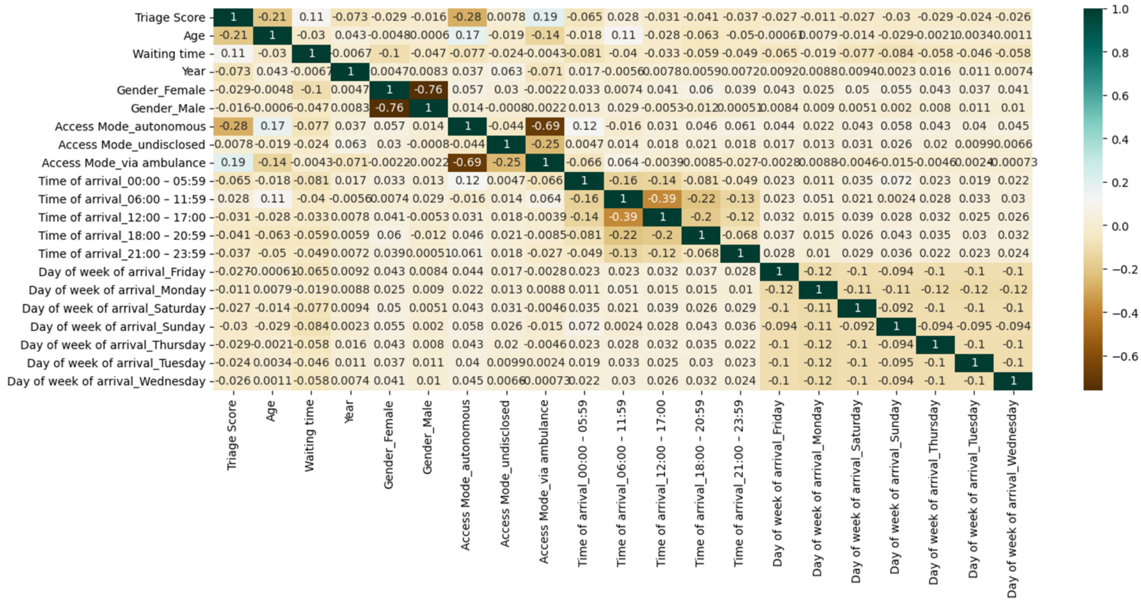

Fig.3. Correlation values between variables.

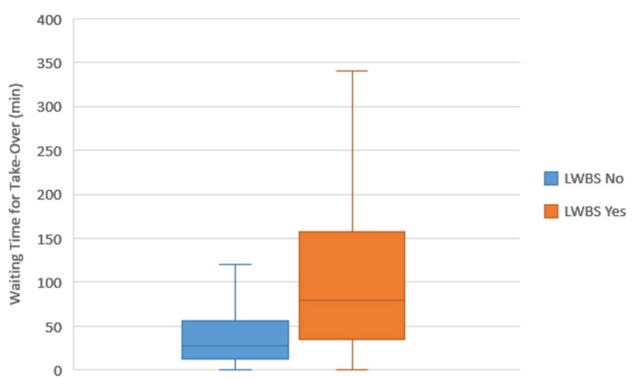

Fig.4. Box plots showing the difference in waiting times for LWBS and non-LWBS patients without outliers.

LWBS / Waiting time for take-over (min)	Mean	Std	Min	25%	50%	75%	Max
No	46.33	74.94	0.00	13.00	27.00	56.00	1439.00
Yes	140.05	209.36	0.00	35.00	80.00	157.00	1439.00

Table 4. Study of waiting time in the two groups.

Algorithms	Best hyperparameters
RF	{'max_depth': 9, 'n_estimators': 20}
NB	{'var_smoothing': 3.5e^{-6}}
LR	{'C': 0.001, 'penalty': 'l2'}
DT	{'max_depth': 17}

Table 5. Best Hyperparameters.

Algorithms	Accuracy	Balanced accuracy	p-value	AUC	Class	F-measure	Precision	Recall
RF	0.87	0.87	< 0.001	0.95	0	0.88	0.85	0.91
					1	0.87	0.90	0.84
NB	0.77	0.76	< 0.001	0.92	0	0.72	0.90	0.60
					1	0.80	0.70	0.93
LR	0.93	0.93	< 0.001	0.96	0	0.94	0.89	0.99
					1	0.93	0.99	0.88
DT	0.94	0.94	< 0.001	0.97	0	0.94	0.91	0.97
					1	0.93	0.97	0.90

Table 6. Evaluation metrics of the ML algorithms.

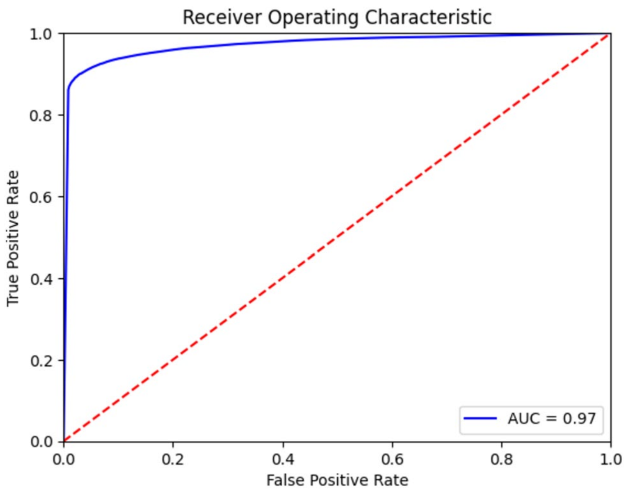

Fig.5. ROC curve.

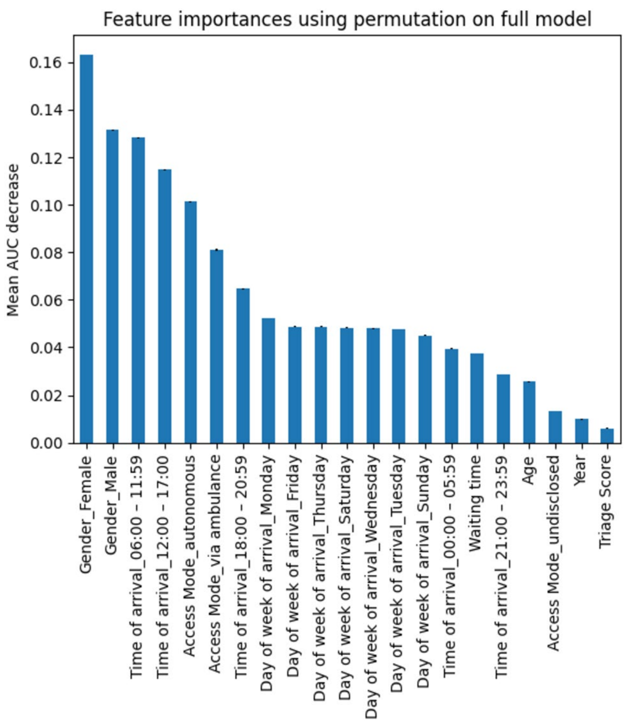

Fig.6. Feature importance.

	coef	Std error	[0.025	0.975]	p-value
Intercept	−7.466	0.583	−266.43	236.87	1.000
Gender					
Female (1)	−2.485	0.006	−24.563	12.831	**0.000**
Male (1)	−2.534	1.50e-5	−4.609	2.479	**0.000**
Triage					
Triage 2	−0.106	0.028	−49.895	56.107	**0.000**
Triage 3	0.764	0.118	−138.439	68.318	1.000
Triage 4	0.001	0.067	−88.186	108.184	1.000
Age	0.065	0.007	−1.630	2.792	**0.000**
Access mode					
Autonomous (1)	4.999	3.64e-5	4.972	5.025	**0.000**
Undisclosed (1)	4.998	0.0002	−0.155	58.790	**0.000**
Ambulance (1)	4.999	1.47e-5	4.988	10.012	**0.000**
Time of arrival					
00:00–05:59 (1)	0.0001	6.72e-5	−0.051	0.050	**0.000**
06:00–11:59 (1)	−6.75e-6	2.20e-5	−0.023	0.019	1.000
12:00–17:59 (1)	−7.20e-6	2.39e-5	−0.023	0.020	1.000
18:00–20:59 (1)	−2.53e-6	4.20e-5	−5.025	5.035	**0.000**
21:00–23:59 (1)	−8.74e-5	9.24e-5	−3.988	0.082	1.000
Day of week of arrival					
Monday (1)	−5.44e-5	3.69e-5	−0.029	0.029	1.000
Tuesday (1)	−1.93e-5	4.10e-5	−1.017	0.036	1.000
Wednesday (1)	−4.27e-6	4.14e-5	−0.038	0.038	1.000
Thursday (1)	−6.97e-6	4.14e-5	−0.047	0.036	1.000
Friday (1)	8.05e-6	4.10e-5	−5.042	0.037	1.000
Saturday (1)	3.85e-5	4.13e-5	−0.035	3.396	1.000
Sunday (1)	6.37e-6	4.49e-5	−0.037	0.036	1.000
Waiting Time	0.029	0.004	−0.462	0.282	**0.000**
Year					
2015	0.052	0.011	−8.898	9.332	1.000
2016	0.054	0.011	−9.189	9.056	1.000
2017	0.053	0.011	−8.758	9.227	1.000
2018	0.052	0.011	−8.792	9.014	1.000
2019	0.052	0.011	−8.754	8.899	1.000
2020	0.036	0.010	−8.233	8.408	1.000
2021	0.002	0.011	−9.107	8.892	1.000

Table 7. Results of the firth logistic regression. Significant values are in bold.

best is in line with what is available in the literature[52], compared to others such as Support Vector Machine and Naïve Bayes that perform worse on unbalanced datasets[53]. The possibility of predicting LWBS in advance with a tolerable error margin supports cost analysis of implemented procedures. Another highlight is that our classification model considered both modifiable and unmodifiable (i.e., sex, age) risk factors for LWBS.

Besides, we believe that analyzing specific aspects of complex health processes enhances overall service perception. Specifically, studying LWBS and demonstrating that length of stay influences the decision to leave can incrementally improve ED efficiency.

Following this logic, one of the main outcomes is that, after a statistical analysis, it was possible to deduce that the waiting time for take-over significantly influence the probability of LBWS. This result is in line with what Rathlev et al.[16] showed in their study. It is actually evident that ED disorganization leads to an increase in waiting times and influences the decision of patients that abandon ED. ED disorganization leads to increased wait times and higher LWBS rates, suggesting that inadequate support staff and resources contribute to inefficiency.

If reducing patient flow below a certain threshold is impossible, increasing available ED resources (e.g., personnel, space, hospital beds) might be necessary. However, this solution could be costly. Alternatively, predicting maximum waiting times in relation to patient flow could enable flexible resource allocation and shift staggering for more concrete solutions.

Direct comparisons with other studies were challenging due to differing contexts, collected variables, and models. A limitation of the work relates precisely to the dataset with limited features due to the small number of variables collected when accessing the ED, which could result in a statistical selection bias. A further problem is related to the bias of outliers, as no actions have been put in place to manage them. The anonymization of the dataset, moreover, did not allow us to verify the presence of repeated accesses to verify the consequences of the phenomenon investigated. Another consideration must be made about the forecasting models implemented, a new element in our work. The good performance of the models remains tied to the use of SMOTE techniques and not on real data. Moreover, knowing a priori the risk of abandonment for each patient makes it possible not only to manage health priorities but also to treat according to waiting time. This could cause health disparities or put in place actions to counteract it.

Furthermore, it may be useful to make comparisons between EDs with similar characteristics of the patient and the hospital, confirming the deductions presented in this paper. It is also probable that, despite the large data set available to us, the exact knowledge of the daily number and the shifts of doctors, nurses, technicians, and other support staff, and including some of this information in our forecasting model would enhance the prediction power. Future developments of this study will also contemplate a higher number of predictor variables to obtain a more accurate prediction model.

Conclusion
In this study, data from patients registered at the "San Giovanni di Dio e Ruggi d'Aragona" University Hospital's ED between 2014 and 2021 were evaluated. Differently from other studies, we utilized a large dataset to develop a robust prediction model based on four different ML algorithms. Over the six-years period considered, the achieved prediction model has proved to be performing in the estimation of the abandonment rates. Additionally, being able to correlate this phenomenon with specific, easily identifiable parameters can help quickly identify bottlenecks in ED organization. A statistical analysis of the collected parameters revealed that the waiting time is the central issue for LWBS.

Data availability
The datasets generated and/or analyzed during the current study are not publicly available for privacy reasons but could be made available from the corresponding author on reasonable request.

Received: 17 November 2023; Accepted: 19 August 2024
Published online: 22 August 2024

References
1. Asaro, P. V., Lewis, L. M. & Boxerman, S. B. Emergency department overcrowding: analysis of the factors of renege rate. *Acad. Emerg. Med.* **14**, 157–162. https://doi.org/10.1197/j.aem.2006.08.011 (2007).
2. Weiss, S. J. *et al.* Relationship between the National ED overcrowding scale and the number of patients who leave without being seen in an academic. *Journal* https://doi.org/10.1016/j.ajem.2005.02.034 (2005).
3. N. Gilboy, P. Tanabe, D. Travers. Emergency Severity Index, Version 4 Implementation Handbook. Rockville: Agency for Healthcare Research and Quality, 2005.
4. Canadian Association of Emergency Physicians. Canada's ERs: We can't afford to wait any longer. <http://www.stopthewaiting.ca> (Version current at June 26, 2007)
5. Ding, R. *et al.* Patients who leave without being seen: Their characteristics and history of emergency department use. *Ann. Emerg. Med.* **48**, 686–693. https://doi.org/10.1016/j.annemergmed.2006.05.022 (2006).
6. Baker, D. W., Steven, C. D. & Brook, R. H. Patient who leave a public hospital ED without being seen causes and consequences. *JAMA* https://doi.org/10.1001/jama.1991.03470080055029 (1991).
7. Rowe, B. H. *et al.* Characteristics of patients who leave emergency departments without being seen. *Acad. Emerg. Med.* **13**, 848–852. https://doi.org/10.1197/j.aem.2006.01.028 (2006).
8. Mekhaldi, R. N., Caulier, P., Chaabane, S., Chraibi, A. & Piechowiak, S. Using machine learning models to predict the length of stay in a hospital setting. In *Trends and innovations in information systems and technologies* (eds Rocha, Á. *et al.*) (Springer, Cham, 2020).
9. Scala, A., Loperto, I., Triassi, M. & Improta, G. Risk factors analysis of surgical infection using artificial intelligence: A single center study. *Int. J. Environ. Res. Public Health* **19**(16), 10021 (2022).
10. Han, J. H. *et al.* The effect of physician triage on emergency department length of stay. *J. Emerg. Med.* **39**(2), 227–233 (2010).
11. Nayeri, N. D. & Aghajani, M. Patients' privacy and satisfaction in the emergency department: A descriptive analytical study. *Nursing Ethics* **17**(2), 167–177 (2010).
12. Rowe, B. H., Channan, P., Bullar, M., Alibha, A. & Saunders, D. Reasons why patients leave without being seen from the emergency department. *Acad. Emerg. Med.* **10**, 513 (2003).
13. Hsia, R. Y. *et al.* Hospital determinants of emergency department left without being seen rates. *Ann. Emerg. Med.* **58**, 24–32. https://doi.org/10.1016/j.annemergmed.2011.01.009 (2011).
14. Improta, G., Colella, Y., Rossi, G., Borrelli, A., Russo, G. & Triassi, M. Use of machine learning to predict abandonment rates in an emergency department. In Proceedings of the 2021 10th International Conference on Bioinformatics and Biomedical Science 2021 Oct 29 (pp. 153–156).
15. Bair, A. E., Song, W. T., Chen, Y. & Morris, B. A. The Impact of Inpatient boarding on ED efficiency a discrete event simulation study. *J. Med. Sys.* https://doi.org/10.1007/s10916-009-9307-4 (2010).
16. Rathlev, N. K. *et al.* Patient characteristics and clinical process predictors of patients leaving without being seen from the emergency department. *West J. Emerg. Med.* **21**, 1218–1226. https://doi.org/10.5811/westjem.2020.6.47084 (2020).
17. Pham, J. C., Ho, G. K., Hill, P. M., McCarthy, M. L. & Pronovost, P. J. National study of patient, visit, and hospital characteristics associated with leaving an emergency department without being seen: Predicting LWBS. *Acad. Emerg. Med.* **16**, 949–955. https://doi.org/10.1111/j.1553-2712.2009.00515.x (2009).
18. Tropea, J. *et al.* Patient who leave without being seen in emergency departments: An analysis of predictive factors and outcomes. *Acad. Emerg. Med.* **19**, 439–447. https://doi.org/10.1111/j.1553-2712.2012.01327.x (2012).
19. Montella, E., Ferraro, A., Sperlì, G., Triassi, M., Santini, S., & Improta, G. Predictive analysis of healthcare-associated blood stream infections in the neonatal intensive care unit using artificial intelligence: A single center study. *Int. J. Environ. Res. Public Health* **19**(5), 2498 (2022).
20. Monzon, J., Friedman, S. M., Clarke, C. & Arenovich, T. Patients who leave the emergency department without being seen by a physician: A control-matched study. *Can. J. Emerg. Med.* **7**, 107–113. https://doi.org/10.1017/s1481803500013063 (2005).
21. Sheraton, M., Gooch, C. & Kashyap, R. Patients leaving without being seen from the emergency department: A prediction model using machine learning on a nationwide database. *JACEP* **1**, 1684–1690. https://doi.org/10.1002/emp2.12266 (2020).

22. Improta, G., Borrelli, A. & Triassi, M. Machine learning and lean six sigma to assess how COVID-19 has changed the patient management of the complex operative unit of neurology and stroke unit: A single center study. *Int. J. Environ. Res. Public Health* **19**(9), 5215 (2022).
23. Bacchi, S. *et al.* Machine learning the prediction of medical inpatient length of stay. *Intern. Med. J.* https://doi.org/10.1111/imj.14962 (2021).
24. Rosa, A. *et al*. Lean Six Sigma to reduce the acute myocardial infarction mortality rate: A single center study. *TQM Journal* **35**(9), 25–41 (2023).
25. Improta, G. *et al*. Five years of Lean Six Sigma in the healthcare: An overview of main tools and results. *Convegno Nazionale di Bioingegneria* 414–417 (2020).
26. Foni, N. O. *et al.* Clinical pathway improves medical practice in total knee arthroplasty. *PLoS ONE* **15**(5), e0232881. https://doi.org/10.1371/journal.pone.0232881 (2020).
27. Daghistani, T. & Alshammari, R. Comparison of statistical logistic regression and randomforest machine learning techniques in predicting diabetes. *JAIT* **11**(2), 78–83. https://doi.org/10.12720/jait.11.2.78-83 (2020).
28. Burlacu, A. *et al.* Using artificial intelligence resources in dialysis and kidney transplant patients: A literature review. *BioMed Res. Int.* https://doi.org/10.1155/2020/9867872 (2020).
29. Improta, G., Simone, T., & Bracale, M. HTA (Health Technology Assessment): a means to reach governance goals and to guide health politics on the topic of clinical Risk management. In World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/12 General Subjects (pp. 166–169) (Springer, Berlin, Heidelberg, 2009).
30. Wang, C., Pan, X., Ye, L., Zhuang, W. & Ma, F. Forecast of hospitalization costs of child patients based on machine learning methods and multiple classification. *JAIT* **9**(4), 89–96. https://doi.org/10.12720/jait.9.4.89-96 (2018).
31. Santini, S. *et al.* Using fuzzy logic for improving clinical daily-care of β-thalassemia patients. In *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE)* (pp. 1–6) (IEEE, 2017).
32. Polevoi, S. K., Quinn, J. V. & Kramer, N. R. Factors associated with patients who leave without being seen. *Acad. Emerg. Med.* **12**, 232–236. https://doi.org/10.1197/j.aem.2004.10.029 (2005).
33. Crilly, J., Bost, N., Thalib, L., Timms, J. & Gleeson, H. Patients who present to the emergency department and leave without being seen: Prevalence, predictors and outcomes. *Eur. J. Emerg. Med. Off. J. Eur. Soc. Emerg. Med.* **20**, 248–255. https://doi.org/10.1097/MEJ.0b013e328356fa0e (2013).
34. Ghafouri, S. M. M. S., & Haji, B. (2019, January). Utilizing a Simulation Approach for Analysis of Patient Flow in the Emergency Department: A Case Study. In 2019 15th Iran International Industrial Engineering Conference (IIIEC) (pp. 151–157). IEEE.
35. Vanbrabant, L., Braekers, K., Ramaekers, K. & Van Nieuwenhuyse, I. Simulation of emergency department operations: A comprehensive review of KPIs and operational improvements. *Computers Ind. Eng.* **131**, 356–381 (2019).
36. Nahhas, A., Awaldi, A. & Reggelin, T. Simulation and the emergency department overcrowding problem. *Procedia Eng.* **178**, 368–376 (2017).
37. Di Laura, D. *et al*. Efficiency measures of emergency departments: An Italian systematic literature review. *BMJ Open Qual.* https://doi.org/10.1136/bmjoq-2020-001058 (2021).
38. Fragapane, G. I., Zhang, C., Sgarbossa, F. & Strandhagen, J. O. An agent-based simulation approach to model hospital logistics. *Int. J. Simul. Model* **18**(4), 654–665 (2019).
39. Ponsiglione, A. M. *et al.* Predictive models for studying emergency department abandonment rates: a bicentric study. In *International Symposium on Biomedical and Computational Biology* (pp. 445–452). (Cham: Springer International Publishing, 2022).
40. Ortíz-Barrios, M. A. & Alfaro-Saíz, J. J. Methodological approaches to support process improvement in emergency departments: A systematic review. *Int. J. Environ. Res. Public Health* **17**(8), 2664 (2020).
41. Ferraro, A. *et al.* Implementation of lean practices to reduce healthcare associated infections. *Int. J. Healthc. Technol. Manage.* **18**, 51. https://doi.org/10.1504/IJHTM.2020.10039887 (2020).
42. G. Improta, P. Natale, L. C. Santillo, M. Triassi. Health worker monitoring: Kalman-based software design for fault isolation in human breathing. EMSS 2014 Proceedings.
43. Cesarelli, G. *et al. DMAIC Approach for the Reduction of Healthcare-Associated Infections in the Neonatal Intensive Care Unit of the University Hospital of Naples '"Federico II"' In European Medical and Biological Engineering Conference* (Springer, 2020).
44. Scala, A., Borrelli, A. & Improta, G. Predictive analysis of lower limb fractures in the orthopedic complex operative unit using artificial intelligence: The case study of AOU Ruggi. *Sci. Rep.* **12**(1), 22153 (2022).
45. Scala, A., Trunfio, T. A. & Improta, G. Classification and regression model to manage the hospitalization for laparoscopic cholecystectomy. *Sci. Rep.* **13**(1), 14700 (2023).
46. Ponsiglione, A. M., Trunfio, T. A., Amato, F., & Improta, G. Predictive analysis of hospital stay after caesarean section a single-center study. *Bioengineering* **10**(4), 440 (2023).
47. Kessels, R., Jones, B. & Goos, P. Using Firth's method for model estimation and market segmentation based on choice data. *J. Choice Modelling* **31**, 1–21 (2019).
48. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825–2830, 2011.
49. Cross Validation, Scikit-learn Python Library. Retrieved June 08, 2024, from https://scikit-learn.org/stable/modules/cross_validation.html
50. GridSearchCV, Scikit-learn Python Library. Retrieved June 08, 2024, from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
51. SMOTE, Over-sampling Methods, Imbalanced-learn Python Library. Retrieved June 08, 2024, from https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html
52. Rodríguez, J. J., Díez-Pastor, J. F. & García-Osorio, C. Ensembles of decision trees for imbalanced data. In *International Workshop On Multiple Classifier Systems Berlin* (eds Sansone, C. *et al.*) (Springer Berlin Heidelberg, 2011).
53. Thölke, P. *et al.* Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage.* **277**, 120253 (2023).
54. Ghaffari, S., Saleh, E., Forsyth, D. and Wang, Y.X., 2021. On the importance of firth bias reduction in few-shot classification. arXiv preprint arXiv:2110.02529.
55. Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **14**(1), 6086 (2024).
56. Japkowicz, N. & Shah, M. Performance evaluation in machine learning. *Mach. Learn. Radiat. Oncol. Theory Appl.* https://doi.org/10.1007/978-3-319-18305-3_4 (2015).
57. Yousefi, M. *et al.* Simulating the behavior of patients who leave a public hospital emergency department without being seen by a physician: A cellular automaton and agent-based framework. *Braz. J. Med. Biol. Res.* **51**, e6961 (2018).

Author contributions

Conceptualization, A.F. and G.I.; methodology, M.R.M. and T.A.T.; validation, A.M.P. and T.A.T.: formal analysis, M.R.M. and T.A.T.; investigation, M.R.M. and T.A.T.; resources, A.F. and G.I.; data curation, M.R.M. and T.A.T.; writing—original draft preparation, M.R.M. and T.A.T; writing—review and editing, A.M.P., F.A. and

G.I.; visualization, A.M.P. and T.A.T; supervision, A.F. and G.I.; project administration, G.I.. All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.A.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.