



Title	Development of a low dimensional model to predict admissions from triage at a pediatric emergency department
Authors(s)	Leonard, Fiona, Gilligan, John, Barrett, Michael
Publication date	2022-08
Publication information	Leonard, Fiona, John Gilligan, and Michael Barrett. "Development of a Low dimensional Model to Predict Admissions from Triage at a Pediatric Emergency Department." Wiley, August 2022. https://doi.org/10.1002/emp2.12779 .
Publisher	Wiley
Item record/more information	http://hdl.handle.net/10197/27201
Publisher's version (DOI)	10.1002/emp2.12779

Downloaded 2026-05-02 01:17:49

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

ORIGINAL RESEARCH

Pediatrics

Development of a low-dimensional model to predict admissions from triage at a pediatric emergency department

Fiona Leonard MSc¹ | John Gilligan PhD² | Michael J. Barrett MB^{3,4}¹Business Intelligence Unit, Children's Health Ireland at Crumlin, Dublin, Ireland²School of Computer Science, Technological University Dublin, Dublin, Ireland³Department of Paediatric Emergency Medicine, Children's Health Ireland at Crumlin, Dublin, Ireland⁴Women's and Children's Health, School of Medicine, University College Dublin, Dublin, Ireland**Correspondence**Michael J. Barrett, MB, Department of Emergency Medicine, Children's Health Ireland at Crumlin, Dublin, Leinster, Ireland.
Email: michael.barrett@ucd.ie**Meetings:** Oral Presentation *Prediction Modelling of Paediatric Emergency Department Hospital Admissions from Triage*, The 20th International Conference on Emergency Medicine (ICEM21), Virtual (June 8–12, 2021), United Arab Emirates, 11th June 2021.**Funding and support:** By JACEP Open policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist. No grant was secured for this study.**Abstract****Objectives:** This study aims to develop and internally validate a low-dimensional model to predict outcomes (admission or discharge) using commonly entered data up to the post-triage process to improve patient flow in the pediatric emergency department (ED). In hospital settings where electronic data are limited, a low-dimensional model with fewer variables may be easier to implement.**Methods:** This prognostic study included ED attendances in 2017 and 2018. The Cross Industry Standard Process for Data Mining methodology was followed. Eligibility criteria was applied to the data set, splitting into 70% train and 30% test. Sampling techniques were compared. Gradient boosting machine (GBM), logistic regression, and naïve Bayes models were created. Variables of importance were obtained from the model with the highest area under the curve (AUC) and used to create a low-dimensional model.**Results:** Eligible attendances totaled 72,229 (15% admission rate). The AUC was 0.853 (95% confidence interval [CI], 0.846–0.859) for GBM, 0.845 (95% CI, 0.838–0.852) for logistic regression and 0.813 (95% CI, 0.806–0.821) for naïve Bayes. Important predictors in the GBM model used to create a low-dimensional model were presenting complaint, triage category, referral source, registration month, location type (resuscitation/other), distance traveled, admission history, and weekday (AUC 0.835 [95% CI, 0.829–0.842]).**Conclusions:** Admission and discharge probability can be predicted early in a pediatric ED using 8 variables. Future work could analyze the false positives and false negatives to gain an understanding of the implementation of these predictions.

1 | INTRODUCTION

1.1 | Background and importance

Emergency department (ED) overcrowding is a global health issue. Rising incidences of hospital overcrowding have led to an increase in

studies attempting to tackle the problem by early prediction of ED admissions using routinely collected data.^{1–3} Sinclair⁴ proposes the most influential factor contributing to overcrowding in the pediatric ED is the presence of hospital admitted patients (boarders). Inpatient boarders in the ED treatment area increase the wait time for other ED patients and can also cause less acute patients to leave.⁵ The implementation of a predictive analytics tool centered on potential admission predictions from the ED and rapidly forecasting patient

Supervising Editor: Katherine Edmunds, MD.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. JACEP Open published by Wiley Periodicals LLC on behalf of American College of Emergency Physicians.

disposition post triage may alleviate ED overcrowding by directly improving ED patient flow.

Early studies on predicting admissions from the ED focused on clinical judgment based prediction⁶⁻⁹ and comparing this judgment with machine learning methods.¹⁰ More recent research has suggested a combined approach using both machine learning and clinical judgment would achieve the best results.¹¹ Other studies centered on specific patient cohorts,¹²⁻¹⁴ such as acute bronchiolitis, progressive modeling approaches were used,¹⁵ and others used natural language processing to create predictors.¹⁶⁻¹⁸ Later studies derived scores from logistic regression (LR) models to determine risk of admission.^{2,3}

The implementation of models to predict admissions – discharges from the ED using electronic health record data will depend on the level of information technology maturity specific to the technology, people, and processes within each hospital environment. A low-dimensional model with fewer variables may be simpler to implement in settings that have limited data recorded in electronic format.

1.2 | Goals of this investigation

The primary objective of this study is to develop and internally validate a low-dimensional predictive model from a pediatric ED based on a limited data set of input variables. We define a low-dimensional model as including as few variables as possible while maintaining a good (comparable to previous studies) discrimination measure (area under the curve [AUC]). A low-dimensional model will be developed, composed of the most important predictors from the initial models (using all variables in the data set) to further demonstrate what can be achieved in countries without a robust electronic health care record system. This would highlight the important variables to focus on for the data collection process. A model composed of data generated at triage will provide early notice of ED patients requiring hospital admission and those to be fast tracked for discharge. The initial models will be used to compare different sampling strategies and machine learning algorithms, with the best performing model being selected to develop the final low-dimensional model.

2 | METHODS

2.1 | Study setting and design

ED data were analyzed from 1 acute tertiary pediatric multiuniversity affiliated teaching hospital in the Republic of Ireland, which also serves as a secondary care facility for the regional pediatric population, with a yearly census of approximately 39,000 visits. This retrospective study was approved by the Hospital Research and Ethics Committee of Children's Health Ireland at Crumlin (formerly Our Lady's Children's Hospital, Crumlin) (GEN/693/18).

The data mining methodology, Cross Industry Standard Process for Data Mining (CRISP-DM) guided the study; consisting of business understanding, data understanding, data preparation, modeling, evalu-

The Bottom Line

Overcrowding in pediatric emergency departments is a problem, and early anticipation of admissions and discharges improves patient flow. This study predicted dispositions using readily available data obtained at triage, creating a model for use in healthcare systems with minimal electronic health resources.

ation, and deployment.¹⁹ All data extraction and transformation were carried out in Microsoft SQL Server Management Studio. Data understanding, model development, and evaluation were performed in R Studio Version 1.1.456. The reporting guidelines for this prognostic study, as set out by Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD),²⁰ were followed. A protocol was developed and published for this study.²¹

The performance of models developed using 3 different machine learning algorithms and various sampling strategies were compared. The top variables of importance were selected from the model with the highest AUC and used to create an additional low-dimensional model, with a greater focus on generalizability.

2.2 | Data collection and transformation

Attendances to the ED from 2017 to 2018, comprising routinely captured data entered up to post-triage (refers to the end of the triage process where the patient is assigned to the next physical space [ED location] and the type of clinician who will see the patient [clinician type]) were included. Two years of data were selected to provide a robust representation of variables and incorporate seasonal changes. Based on previous studies we excluded patients over 18 years of age,²² visits with a did not wait outcome (includes left before being seen and left before completion of treatment) as their outcome was non-deterministic should the patient have stayed,^{1,3,12,22-25} and patients with missing data (sex, triage category,²² and health care record number).²⁶ As outlined in our protocol²¹ missing data were analyzed to assess the most appropriate method to address; listwise deletion was selected due to the small percentage and the assumption that data was missing at random. Patients returning for specimen collection or day case management were also excluded as their inclusion could be a potential source of bias.

The data set creation, outcome variable (admission or discharge), and initial variable selection are outlined in our protocol²¹ and consisted of 34 variables extracted from 3 separate databases. The outcome of "admission" included attendances with a discharge outcome of "admission," "transferred to another hospital for admission" and "died in department." The data set represents demographics, registration details, triage assessment (triage category was allocated using the Irish Children's Triage System²⁷), hospital usage, past medical

TABLE 1 Feature engineering tasks performed to transform variables

Variable	Transformation description
Disposition (outcome variable)	Transformed to 0 or 1, assigned a value of 1 for discharge outcome of “admission” or “transferred to another hospital for admission” ^{22,28,29} or “died in department.” ³ All other discharge outcomes defaulted to 0.
Age	Split into 5 groups consisting of neonate (0–28 days), infant (28 days–1 year), preschool (2–5 years), school age (6–12 years), and adolescent (13–18 years).
Registration hour	Grouped into 4-hour intervals, ^{25,30} 0–4, 4–8, 8–12, 12–16, 16–20, and 20–24.
Arrival mode	Recoded to “Ambulance,” “Private Transport,” and “Other.”
Referral source	Grouped into “Self,” “Other Hospitals,” “General Practitioner,” “Clinic,” and “Other.”
Triage category	To mitigate against possible quasi-complete separation ³¹ for triage 1 (most acute) and 5 we grouped triage into 3 categories: 1–2, 3, and 4–5
Presenting complaint	To lower the cardinality that may impact model performance, ³² presenting complaint categories were reduced based on clinician expertise, from 177 separate categories down to 58.
Complex chronic conditions	Eleven new binary variables based on pediatric complex chronic conditions. ³³ These were derived using the ICD-10-AM diagnosis from all previous admissions.
Diagnosis related groups	Based on specific cohorts of patients frequenting this facility, the last 3 years admission diagnosis related groups were included as binary variables. Blood immunology and digestive system groups were created.
Distance travelled	Calculated using the GPS coordinates from the patient’s address to the hospital site and grouped into kilometer groupings of 0–2, 2–4, 4–6, 6–10, 10–20, 20–40, 40–60, 60–100, and 100+ kilometers.
Emergency department location	First location the patient assigned to at the end of triage which was grouped into “Resuscitation” and “Other.”
Clinician type	Type of clinician patient assigned to post-triage grouped into “Advanced Nurse Practitioner” and “Other.”
Infection control alert	Encoded to 1 if value is present and 0 if absent.
Number of emergency department attendances in the last year	Recoded into groups of 0, 1, 2, 3–4, and 5+ attendances.
Number of admissions in the last year	Recoded into groups of 0, 1, 2, 3, 4, and 5+ admissions.
Admitted in last 7 days	Encoded to 1 if value is present and 0 if absent.
Admitted in last 30 days	Encoded to 1 if value is present and 0 if absent.
Admitted in last 3 years	Encoded to 1 if value is present and 0 if absent.
Any previous admission	Encoded to 1 if value is present and 0 if absent.

Abbreviation: ICD-10-AM, International Statistical Classification of Diseases, Tenth Revision, Australian Modification

history, first ED location allocated to post-triage (grouped into patients going to resuscitation and “other” for all other ED spaces), and first clinician type seen. The raw data set was analyzed using location, dispersion, and shape statistics for continuous variables and stacked bar charts for categorical variables. After the data understanding process, the final data set was formalized through feature engineering tasks (Table 1). All continuous variables were converted to categorical after an initial evaluation through a LR model using their original state, with the decision to transform based on model performance.

2.3 | Statistical analyses

Descriptive statistics consisting of the number of visits and percentage were used to analyze the data. χ^2 statistics were used to assess independence of each variable with respect to the outcome. Following previous research^{1,12,22} the data were partitioned into approximately

70% training and 30% testing using random sampling, while maintaining the relative ratios of the values in the variables. The training set was used in model creation and the test set was used for internal validation, providing an unbiased evaluation of the resulting models.

The 3 machine learning algorithms; LR,^{1,23,26,28,31,34} naïve Bayes (NB),^{15,34} and gradient boosting machine (GBM)^{22,35,36} were selected based on their use in previous research. Sampling approaches and detailed configuration for the machine learning algorithms can be found in Table SE1 (Appendix E1).

Evaluation of model performance using the test set was measured primarily using AUC with 95% confidence intervals (CIs) using the DeLong method. Sensitivity, specificity, accuracy, positive prediction value (PPV), and negative prediction value (NPV) were used as the secondary measurements with 95% CI. To compare the secondary measurements across the different machine learning algorithms, the method of threshold-moving by fixing the specificity at 90% was applied. The variables of importance were obtained from the model

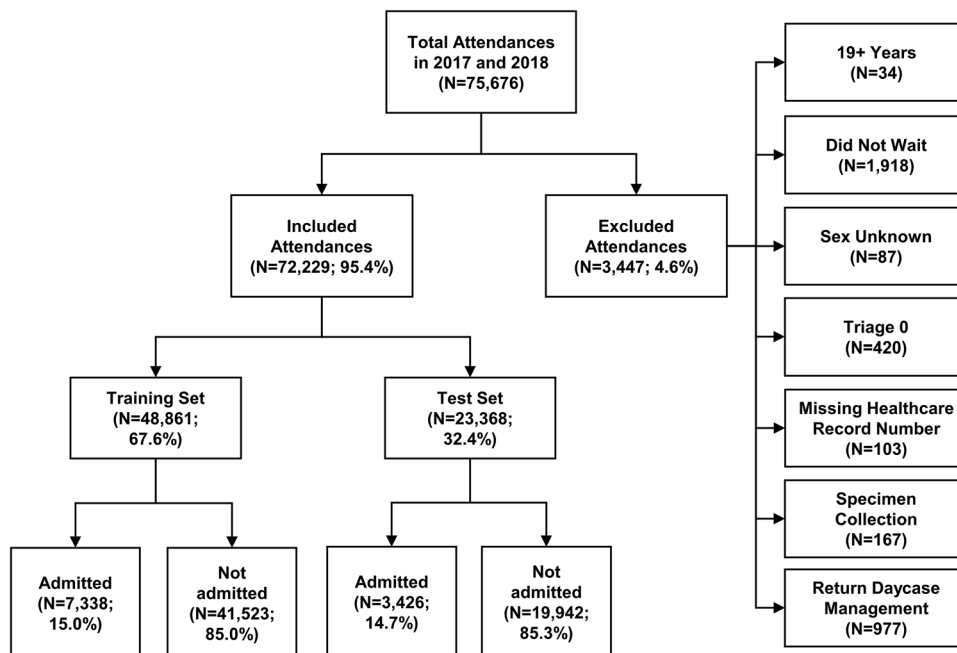


FIGURE 1 Total visits to the emergency department in 2017 and 2018, summarizing exclusions and data partitioning between training and test set

with the highest AUC, for the GBM model this was achieved by using relative influence, based on the associated average decrease in mean squared error. The top variables of importance from the GBM reference model were used to create an additional low-dimensional model that could potentially be deployed as a decision support tool.³⁰

3 | RESULTS

3.1 | Characteristics of study dataset

The combined ED census for 2017 and 2018 was 75,676. A total of 3447 attendances were excluded from this study. Most patients (66%) who did not wait had left before triage or were low acuity (Category 4–5). Attendances excluded solely due to missing data were 357 (less than 0.5%). The final data set had 72,229 attendances related to 44,944 unique patients. The admission rate at approximately 15% was distributed between the training and test sets, emphasizing the class imbalance between “admission” and “discharged” (Figure 1).

Neonates had the highest rate of admission at 34.4%, but lowest rate of attendances to the ED (1.9%). Most admissions are infants at 35.5%. The admission rate increased according to the triage category, showing a rate of 39.5% for triage 1–2 and down to 5% for triage 4–5. Patients allocated at triage to one of the resuscitation bays showed a high percentage rate of admission at 55.2%. The greater the distance the patient must travel to the ED, the more likely that this will result in an outcome of admission, with admitting rates ranging from 9.9% for 0–2 km up to 29.3% for patients travelling 100 km or more. Most patients arrive using private transport (94.6%), but those arriving by ambulance have a higher admission rate (32.4%). Patients referred

from other hospitals have a greater rate of admission at 47.5%. Interestingly 58.4% of the patient visits had no previous attendance to the ED and 84.5% had no previous admissions, both applicable to within 1 year of the patient’s attendance; 63.2% of the ED visits had no previous admissions. The presenting complaint of “injury” had the highest number of attendances, making up 17.1% of the total, with a 5% admission rate. Although, psychiatric presentations accounted for just 0.5% of the attendances, the admission rate was the highest at 44.1% (Table 2).

The χ^2 test for each variable produced a $P < 0.001$, except for sex, weekday, month, and the Complex Chronic Condition of “Miscellaneous Other.” As a stepwise approach was taken to model creation including these variables had no negative impact on performance, with the Akaike information criterion reducing with the addition of each variable (some of these steps are presented in Table 3). Further descriptive statistics and results of the χ^2 tests for significance are presented in Table SE2 (Appendix E1).

3.2 | Model generation and evaluation

We observed no significant improvement in AUC by applying the additional sampling methods. The GBM model achieved the highest AUC at 0.853 (95% CI, 0.846–0.859) and the highest sensitivity at 56.39 (95% CI, 54.71–58.06), accuracy 85.07 (95% CI, 84.61–85.53), PPV 49.21 (95% CI, 47.63–50.79), and NPV 92.32 (95% CI, 91.93–92.69), making it the best performing model based on AUC alone (Table 4). Results of the application of additional sampling methods can be found in Table SE3 (Appendix E1).

Incidentally, we also looked at the impact of removing patients presenting with a psychiatric complaint from the LR model due to the high

TABLE 2 Descriptive statistics for each variable with respect to the outcome

Variable	Value	Admission		Discharged		P value
		No. visits	% of visits	No. visits	% of visits	
Age						<0.001
	0–28 days	479	4.5%	915	1.5%	
	28 days–23 months	3818	35.5%	16,469	26.8%	
	2–5 years	2567	23.8%	17,830	29.0%	
	6–12 years	2564	23.8%	18,878	30.7%	
	13–18 years	1336	12.4%	7373	12.0%	
Sex						0.036
	Male	5926	55.1%	34,509	56.1%	
	Female	4838	44.9%	26,956	43.9%	
Arrival mode						<.001
	Private transport	9515	88.4%	58,823	95.7%	
	Ambulance	1232	11.4%	2566	4.2%	
	Other	17	0.2%	76	0.1%	
Referral source						<0.001
	Self	6771	62.9%	42,872	69.8%	
	Other hospitals	1072	10.0%	1184	1.9%	
	Swift/other clinic	119	1.1%	1073	1.7%	
	General Practitioner	2701	25.1%	16,019	26.1%	
	Other	101	0.9%	317	0.5%	
Distance traveled						<0.001
	0–2K	723	6.7%	6566	10.7%	
	2–4K	1430	13.3%	11,852	19.3%	
	4–6K	1732	16.1%	9885	16.1%	
	6–10K	1638	15.2%	9480	15.4%	
	10–20K	1904	17.7%	10,223	16.6%	
	20–40K	1489	13.8%	7039	11.5%	
	40–60K	634	5.9%	2696	4.4%	
	60–100K	795	7.4%	2712	4.4%	
	100+K	419	3.9%	1012	1.6%	
Triage category						<0.001
	0–1	5128	47.6%	7869	12.8%	
	3	4093	38.0%	24,495	39.8%	
	4–5	1543	14.3%	29,120	47.4%	
Emergency department location						<0.001
	Resuscitation	2420	22.5%	1968	3.2%	
	Other	8344	77.5%	59,497	96.8%	
Number of attendances in last year						<0.001
	0	6142	57.1%	36,063	58.7%	
	1	2093	19.4%	13,243	21.5%	
	2	1034	9.6%	5708	9.3%	
	3–4	925	8.6%	4206	6.8%	
	5+	570	5.3%	2245	3.7%	

(Continues)

TABLE 2 (Continued)

Variable	Value	Admission		Discharged		P value
		No. visits	% of visits	No. visits	% of visits	
Number of admissions in last year						<0.001
	0	8046	74.7%	53,022	86.3%	
	1	1353	12.6%	5595	9.1%	
	2	553	5.1%	1502	2.4%	
	3	289	2.7%	560	0.9%	
	4	171	1.6%	294	0.5%	
	5+	352	3.3%	492	0.8%	
Admission history						<0.001
	0	5682	52.8%	39,978	65.0%	
	1	5082	47.2%	21,487	35.0%	
Presenting complaint (top 5)						<0.001
	Injury	611	5.7%	11,723	19.1%	
	Vomiting	1079	10.0%	3831	6.2%	
	Difficulty breathing	1101	10.2%	3572	5.8%	
	Abdominal pain	613	5.7%	3599	5.9%	
	Fever	644	6.0%	2916	4.7%	

admission rate. This resulted in an AUC of 0.842 (95% CI, 0.835–0.849) compared to 0.845 (95% CI, 0.838–0.852) for the model that included psychiatric presentations.

Variables representing presenting complaint, triage category, referral source, registration month, first ED location after triage (resuscitation or other), distance traveled, history of any admission, and weekday were the top predictors from our GBM model (Figure 2). The rationale for selecting the top 8 variables was based on the resulting AUC. The top 7 variables, which excluded weekday, resulted in an AUC of 0.832 (95% CI, 0.825–0.838) and the top 9 variables (including registration hour) produced an AUC of 0.833 (95% CI, 0.826–0.840). From the top 8 variables, a low-dimensional GBM model generated an AUC of 0.835 (95% CI, 0.829–0.842) and with a fixed specificity of 90%, produced sensitivity of 52.61 (95% CI, 50.93–54.27), accuracy of 84.58 (95% CI, 84.12–85.03), PPV 47.16 (95% CI, 45.59–48.75), and NPV 91.80 (95% CI, 91.41–92.17).

4 | LIMITATIONS

Although, this research was limited to data from 1 ED of a tertiary standalone pediatric hospital in Ireland, our study shows that the variables of importance consist of data that are routinely collected in most hospitals, therefore increasing model generalizability. Initial generalization in our approach to developing a low-dimensional model was central. We acknowledge data, such as serial age specific vitals, prehospital interventions, pain scores, medications, anthropometrics, and tests ordered could potentially improve model performance, but the objective of this study was to develop a model based on limited electronic

data availability. Outlined in our protocol,²¹ predictor-based limitations include the lack of standardization of presenting complaints, triage categories, and the grouping of ED location into “resuscitation” or “other,” whose inclusion proved significant. Although, a strong predictor in our model, the history of admission may contain past admissions unrelated to the reason for presentation to the ED.

5 | DISCUSSION

This study has created prediction models that are pediatric focused,^{22–24} concentrating on low-dimensional early prediction,^{1,28,37} comparing the performance of different machine learning algorithms^{30,34,35} to identify the optimal model and proposed utility for certain stakeholders (clinicians and bed flow managers). Machine learning algorithms can predict the probability of admission and discharge in the pediatric ED using routinely collected data captured up to the post-triage process. We identified 8 key predictors to create a further low-dimensional prediction model demonstrating what can be achieved with a minimum data set, whose predictive ability was not much different to the model that included 33 variables.

Our GBM reference model with all predictors produced an AUC of 0.853, which outperforms many other studies.^{1,2,34,37} Compared to other pediatric studies, Goto et al²² focused on hospitalization at triage prediction and achieved an AUC of 0.80. Barak-Corren et al²³ used a progressive time approach and achieved an AUC of 0.868 after 10 minutes and 0.913 after 60 minutes; however, the study included data volumes (6009 variables) that may not be available in many hospitals. A lower number of variables may be more readily accessible and to

TABLE 3 Stepwise approach to variable inclusion using logistic regression

Variables	AIC	Delta AIC	AUC (95% CI)
Triage only (1,2,3,4,5)			0.745 (0.738–0.752)
Triage only (grouped 1–2, 3, 4–5)	36,207.7	4895.0	0.740 (0.734–0.747)
+ Age group	35,986.3	4673.6	0.748 (0.739–0.757)
+ Arrival mode	35,871.6	4558.9	0.757 (0.750–0.764)
+ Referral source	33,931.4	2618.7	0.793 (0.785–0.799)
+ Distance in kilometers	33,719.1	2406.4	0.800 (0.793–0.808)
+ Admission history	33,363.2	2050.5	0.805 (0.799–0.812)
+ Reattender (within 7 days)	33,115.3	1802.6	0.813 (0.806–0.820)
+ Presenting complaint	32,223.3	910.6	0.830 (0.823–0.837)
+ Emergency department location (resuscitation or other)	31,694.2	381.5	0.829 (0.822–0.837)
+ Admitted in last 3 years	31,691.0	378.3	0.827 (0.820–0.834)
+ Number of visits in last year	31,667.0	354.3	0.832 (0.826–0.839)
+ Blood immunology group	31,595.2	282.5	0.837 (0.831–0.844)
+ Digestive system group	31,579.5	266.8	0.835 (0.828–0.841)
+ Admitted in last 7 days	31,547.4	234.7	0.834 (0.827–0.841)
+ Admitted in last 30 days	31,529.5	216.8	0.835 (0.828–0.842)
+ Number of admissions in last year	31,475.1	162.4	0.834 (0.827–0.840)
+ Clinician type (ANP or other)	31,413.4	100.7	0.834 (0.828–0.841)
+ Infection control alert	31,409.8	97.1	0.835 (0.828–0.842)
+ Complex chronic conditions	31,379.7	67.0	0.832 (0.825–0.839)
+ Registration hour	31,347.6	34.9	0.838 (0.832–0.845)
+ Sex	31,326.2	13.5	0.834 (0.827–0.841)
+ Weekday	31,320.3	7.6	0.834 (0.827–0.841)
+ Registration month (all variables)	31,312.7	0	0.845 (0.838–0.852)

Note: Delta AIC shows the difference in AIC between the model with the best fit (lowest AIC) and the comparison model.

Abbreviations: AIC, Akaike information criterion; ANP, advanced nurse practitioner; AUC, area under the curve; CI, confidence interval.

TABLE 4 Performance of machine learning algorithm at a fixed specificity of 90%, evaluated using the test set

Machine Learning Algorithm	AUC (95% CI)	% Sensitivity (95% CI)	% Accuracy (95% CI)	% PPV (95% CI)	% NPV (95% CI)
Naïve Bayes	0.812 (0.805–0.819)	46.00 (44.32–47.69)	83.55 (83.07–84.02)	44.15 (42.51–45.79)	90.66 (90.24–91.06)
Logistic regression	0.845 (0.838–0.852)	55.28 (53.60–56.96)	84.91 (84.45–85.37)	48.71 (47.13–50.30)	92.14 (91.75–92.51)
Gradient boosting machine	0.853 (0.846–0.859)	56.39 (54.71–58.06)	85.07 (84.61–85.53)	49.21 (47.63–50.79)	92.32 (91.93–92.69)

Abbreviations: AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

generalize further, we followed the approach by Hong et al³⁰ of identifying the top variables of importance from the best performing model to create a further low-dimensional model.

The low-dimensional model was developed to demonstrate what could be achieved with even fewer variables, the GBM model achieved an AUC of 0.835. Apparent from this model is the exclusion of

patient age, which has been described³⁰ as highly significant in many studies,^{1,3,30} indicating a difference between general (adult and pediatric) and exclusively pediatric settings. Age is a central component to the Irish Children's Triage System,²⁷ and therefore it is still reflected in the model. Also evident is the position of "admission history" as a top predictor, pediatric patient history will be shorter and including an

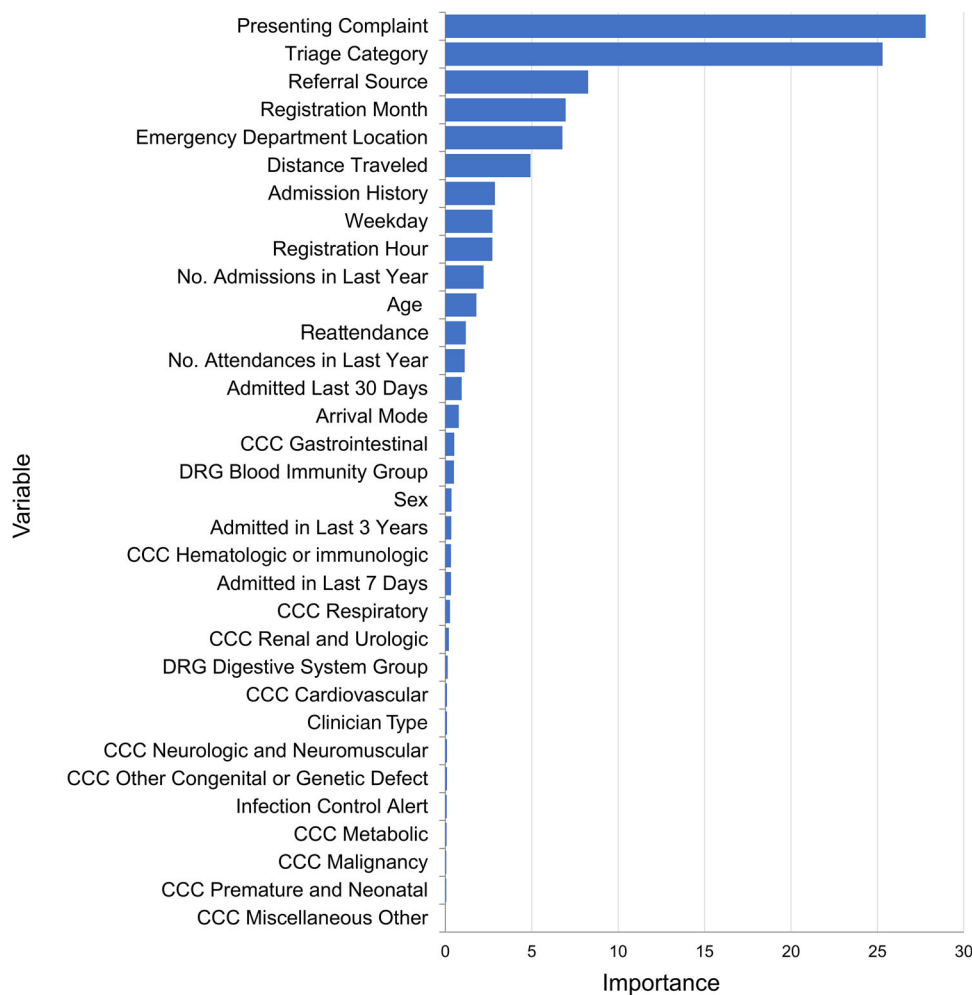


FIGURE 2 Variable importance according to the reference model for gradient boosting machine. Importance measured by the average decrease in mean squared error. Abbreviations: CCC, complex chronic condition; DRG, diagnosis related group; ED, emergency department

indication of any previous admissions proved significant. Other hospitals with limited data may adopt a similar approach of discovering useful variables that may differ from variables/characteristics used in this and previous research to develop a prediction model.

Although the output of this prediction model could be used as a binary score, the output as a probability would have far more use in practice. For ED clinician use, instead of limiting the predictive capabilities to admission or discharge, the probability output could be used for streaming into “assessment for admission,” “fast track for discharge,” and “senior review.”³ Compared to a non-streamed approach, evidence has shown that grouping ED patients into work streams results in a reduction in wait times and a shorter length of stay in the ED, with limited evidence that grouping patients solely based on whether or not they are going to be hospitalized improves ED patient flow.^{38,39} Sun et al³⁷ proposed the model threshold should be adjusted based on the intended use in practice (clinician or bed manager). As our model produces a lower sensitivity, the 8-variable model with the threshold set at 0.5 achieves a specificity of 97.27 and could be used by clinicians to identify patients to “fast track for discharge” or “assessment for admission.” For patients that achieved an inter-

mediate probability, these may be more appropriate for the “senior review” stream as their decision to admit or discharge may be more difficult to assess.³ In the ED, senior clinician review improves patient safety and departmental flow by increasing the accuracy of disposition decisions.^{3,40}

For bed management use, a low threshold of 0.28 on our model produces a sensitivity of 52.61. This would assist bed managers in identifying patients to be admitted, therefore speeding up the advance bed planning process. Summing the individual admission probabilities would provide a total number of beds required, without over inflating. Bed managers might also use this total bed demand to make decisions like expediting inpatient discharges to create capacity.⁴¹

There are many different approaches to deploying the output of the model. Some researchers suggest embedding the prediction results within the electronic medical record,^{1,15,30} which would be useful as additional decision support information in ED patient tracking systems.³⁴ System alerts could be used to prompt triage nurses to trigger a bed request for patients with a high probability of admission, while awaiting clinical decision making.³⁷ A simpler approach would be for the model to power a dynamic dashboard,²³ the output

of our prediction models could be displayed on a real-time dashboard for clinician or bed management use. Many groups have highlighted the benefit of using dashboards in the ED to visualize current status and overcrowding,⁴²⁻⁴⁴ adding an indication of potential admissions and patients to be fast tracked for discharge at an early stage could significantly improve situational awareness expediting patient care. Prototypes demonstrating the intended use are shown in Figures SE1 and SE2 (Appendix E1).

This study has developed a low-dimensional model that predicts admissions and discharges from a pediatric ED using data collected early in the patient's journey. Future work could analyze both the false positives and false negatives to gain an understanding of the implementation of these prediction models. Further research could expand this study beyond 1 hospital site to a network of hospitals. This would further the development of a robust and reliable model capable of positively affecting the patient's journey.

ACKNOWLEDGMENT

Open access funding provided by IREL.

CONFLICTS OF INTEREST

All authors have no conflicts of interest to disclose.

AUTHOR CONTRIBUTIONS

Fiona Leonard, John Gilligan, and Michael J. Barrett conceived the study. Fiona Leonard extracted the data, carried out the analyses (with guidance from John Gilligan and Michael J. Barrett), and prepared and finalized the manuscript. Fiona Leonard, John Gilligan, and Michael J. Barrett have made a substantial contribution to this manuscript, give their final approval of the submitted version, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

- Parker CA, Liu N, Wu SX, et al. Predicting hospital admission at the emergency department triage: a novel prediction model. *Am J Emerg Med*. 2019;37(8):1498-1504. <https://doi.org/10.1016/j.ajem.2018.10.060>
- Dinh MM, Russell SB, Bein KJ, et al. The Sydney Triage to Admission Risk Tool (START) to predict emergency department disposition: a derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. *BMC Emerg Med*. 2016;16(1):46. <https://doi.org/10.1186/s12873-016-0111-4>
- Cameron A, Rodgers K, Ireland A, et al. A simple tool to predict admission at the time of triage. *Emerg Med J*. 2015;32(3):174-179. <https://doi.org/10.1136/emered-2013-203200>
- Sinclair D. Emergency department overcrowding - implications for paediatric emergency medicine. *Paediatr Child Health*. 2007;12(6):491-494. <https://doi.org/10.1093/pch/12.6.491>
- Chan M, Meckler G, Doan Q. Paediatric emergency department overcrowding and adverse patient outcomes. *Paediatr Child Health*. 2017;22(7):377-381. <https://doi.org/10.1093/pch/pxx111>
- Brillman JC, Doezema D, Tandberg D, et al. Triage: limitations in predicting need for emergent care and hospital admission. *Ann Emerg Med*. 1996;27(4):493-500. [https://doi.org/10.1016/S0196-0644\(96\)70240-8](https://doi.org/10.1016/S0196-0644(96)70240-8)
- Levine SD, Colwell CB, Pons PT, et al. How well do paramedics predict admission to the hospital? A prospective study. *J Emerg Med*. 2006;31(1):1-5. <https://doi.org/10.1016/j.jemermed.2005.08.007>
- Beardsell I, Robinson S. Can emergency department nurses performing triage predict the need for admission? *Emerg Med J*. 2011;28(11):959-962. <https://doi.org/10.1136/emj.2010.096362>
- Stover-Baker B, Stahlman B, Pollack M. Triage nurse prediction of hospital admission. *J Emerg Nurs*. 2012;38(3):306-310. <https://doi.org/10.1016/j.jen.2011.10.003>
- Peck JS, Benneyan JC, Nightingale DJ, et al. Predicting emergency department inpatient admissions to improve same-day patient flow. *Acad Emerg Med*. 2012;19(9):E1045-E1054. <https://doi.org/10.1111/j.1553-2712.2012.01435.x>
- Barak-Corren Y, Agarwal I, Michelson KA, et al. Prediction of patient disposition: comparison of computer and human approaches and a proposed synthesis. *J Am Med Inform Assoc*. 2021;28(8):1736-1745. <https://doi.org/10.1093/jamia/ocab076>
- Goto T, Camargo CA, Faridi MK, et al. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am J Emerg Med*. 2018;36(9):1650-1654. <https://doi.org/10.1016/j.ajem.2018.06.062>
- Marlais M, Evans J, Abrahamson E. Clinical predictors of admission in infants with acute bronchiolitis. *Arch Dis Child*. 2011;96(7):648-652. <https://doi.org/10.1136/adc.2010.201079>
- Considine J, Botti M, Thomas S. Early predictors of hospital admission in emergency department patients with chronic obstructive pulmonary disease. *Australas Emerg Nurs J*. 2011;14(3):180-188. <https://doi.org/10.1016/j.aenj.2011.05.004>
- Barak-Corren Y, Israelit SH, Reis BY. Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. *Emerg Med J*. 2017;34(5):308-314. <https://doi.org/10.1136/emered-2014-203819>
- Lucini FR, Fogliatto FS, da Silveira GJC, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. *Int J Med Inform*. 2017;100:1-8. <https://doi.org/10.1016/j.ijmedinf.2017.01.001>
- Zhang X, Kim J, Patzer RE, et al. Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods Inf Med*. 2017;56(05):377-389. <https://doi.org/10.3414/ME17-01-0024>
- Sterling NW, Patzer RE, Di M, et al. Prediction of emergency department patient disposition based on natural language processing of triage notes. *Int J Med Inform*. 2019;129:184-188. <https://doi.org/10.1016/j.ijmedinf.2019.06.008>
- Wirth R, Hipp J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining. 2000;1:29-39. <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>. Accessed October 7, 2018.
- Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. <https://doi.org/10.7326/M14-0697>
- Leonard F, Gilligan J, Barrett MJ. Predicting admissions from a paediatric emergency department - Protocol for developing and validating a low-dimensional machine learning prediction model. *Front Big Data*. 2021;4:643558. <https://doi.org/10.3389/fdata.2021.643558>
- Goto T, Camargo CA, Faridi MK, et al. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open*. 2019;2(1):e186937. <https://doi.org/10.1001/jamanetworkopen.2018.6937>
- Barak-Corren Y, Fine AM, Reis BY. Early prediction model of patient hospitalization from the pediatric emergency department. *Pediatrics*. 2017;139(5):e20162785. <https://doi.org/10.1542/peds.2016-2785>

24. Leegon J, Jones I, Lanaghan K, et al. Predicting hospital admission in a pediatric emergency department using an artificial neural network. *AMIA Annu Symp Proc.* 2006;1004. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839665/>. Accessed October 7, 2018.
25. Leegon J, Jones I, Lanaghan K, et al. Predicting hospital admission for emergency department patients using a Bayesian network. *AMIA Annu Symp Proc.* 2005;1022. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560667/>
26. Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Acad Emerg Med.* 2018;25(12):1463-1470. <https://doi.org/10.1111/acem.13655>
27. ICTS National Steering Group. *National Emergency Medicine Programme Irish Children's Triage System (ICTS)*. <https://healthservice.hse.ie/filelibrary/onmsd/national-emergency-medicine-programme-irish-childrens-triage-system-icts.pdf>. Published September 2016. Accessed September 24, 2020.
28. Lucke JA, de Gelder J, Clarijs F, et al. Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. *Emerg Med J.* 2018;35(1):18-27. <https://doi.org/10.1136/emered-2016-205846>
29. Levin S, Toerper M, Hamrock E, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Ann Emerg Med.* 2018;71(5):565-574.e2. <https://doi.org/10.1016/j.annemergmed.2017.08.005>
30. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE.* 2018;13(7):e0201016. <https://doi.org/10.1371/journal.pone.0201016>. Dong Q, ed.
31. Kraaijvanger N, Rijpsma D, Roovers L, et al. Development and validation of an admission prediction tool for emergency departments in the Netherlands. *Emerg Med J.* 2018;35(8):464-470. <https://doi.org/10.1136/emered-2017-206673>
32. Moeyersoms J, Martens D. Including high-cardinality attributes in predictive models: a case study in churn prediction in the energy sector. *Decis Support Syst.* 2015;72:72-81. <https://doi.org/10.1016/j.dss.2015.02.007>
33. Feudtner C, Feinstein JA, Zhong W, et al. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr.* 2014;14(1):199. <https://doi.org/10.1186/1471-2431-14-199>
34. Rendell K, Koprinska I, Kyme A, et al. The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making. *Emerg Med Australas.* 2019;31(3):429-435. <https://doi.org/10.1111/1742-6723.13199>
35. Graham B, Bond R, Quinn M, et al. Using data mining to predict hospital admissions from the emergency department. *IEEE Access.* 2018;6:10458-10469. <https://doi.org/10.1109/ACCESS.2018.2808843>
36. Araz OM, Olson D, Ramirez-Nafarrate A. Predictive analytics for hospital admissions from the emergency department using triage information. *Int J Prod Econ.* 2019;208:199-207. <https://doi.org/10.1016/j.ijpe.2018.11.024>
37. Sun Y, Heng BH, Tay SY, et al. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med.* 2011;18(8):844-850. <https://doi.org/10.1111/j.1553-2712.2011.01125.x>
38. Jarvis PRE. Improving emergency department patient flow. *Clin Exp Emerg Med.* 2016;3(2):63-68. <https://doi.org/10.15441/ceem.16.127>
39. Oredsson S, Jonsson H, Rognes J, et al. A systematic review of triage-related interventions to improve patient flow in emergency departments. *Scand J Trauma Resusc Emerg Med.* 2011;19(1):43. <https://doi.org/10.1186/1757-7241-19-43>
40. White AL, Armstrong PAR, Thakore S. Impact of senior clinical review on patient disposition from the emergency department. *Emerg Med J.* 2010;27(4):262-265. <https://doi.org/10.1136/emj.2009.077842>
41. Peck JS, Gaehde SA, Nightingale DJ, et al. Generalizability of a simple approach for predicting hospital admission from an emergency department. *Acad Emerg Med.* 2013;20(11):1156-1163. <https://doi.org/10.1111/acem.12244>
42. Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: supporting real-time throughput decision-making. *J Biomed Inform.* 2017;71:211-221. <https://doi.org/10.1016/j.jbi.2017.05.024>
43. Martin N, Bergs J, Eerdeken D, et al. Developing an emergency department crowding dashboard: a design science approach. *Int Emerg Nurs.* 2018;39:68-76. <https://doi.org/10.1016/j.ienj.2017.08.001>
44. Yoo J, Jung KY, Kim T, et al. A real-time autonomous dashboard for the emergency department: 5-Year case study. *JMIR Mhealth Uhealth.* 2018;6(11):e10666. <https://doi.org/10.2196/10666>

AUTHOR BIOGRAPHY



Michael Barrett is a Clinical Associate Professor in the Women's and Children's Health, School of Medicine at University of Dublin and Consultant in Pediatric Emergency Medicine at Children's Health Ireland (CHI), Dublin, Ireland.

SUPPORTING INFORMATION

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

How to cite this article: Leonard F, Gilligan J, Barrett MJ. Development of a low-dimensional model to predict admissions from triage at a pediatric emergency department. *JACEP Open.* 2022;3:e12779. <https://doi.org/10.1002/emp2.12779>