Using Machine Learning to Predict Patients Who Leave Without Being Seen in a Pediatric Emergency Department

by

Julia Sarty

Submitted in partial fulfilment of the requirements

for the degree of Masters of Applied Science

at

Dalhousie University

Halifax, Nova Scotia

April 2023

Dalhousie University is located in Mi'kma'ki, the

ancestral and unceded territory of the Mi'kmaq.

We are all Treaty people.

Table of Contents

List of Tables

List of Figures

Abstract

Background: Patients and their caregivers who seek care in an Emergency Department (ED) may ultimately choose to leave without being seen by a physician. This occurrence is labeled "left without being seen" (LWBS) and can account for up to 15% of all patients who come to an ED. Patients who LWBS do not receive the care they sought in the ED and may experience clinical deterioration related to delayed diagnosis or treatment.

Objective: To describe a LWBS cohort and identify key LWBS attributes in a Canadian pediatric emergency department through thorough machine learning analysis. This prediction is intended to be used in practice to prevent adverse outcomes related to LWBS. This study focuses on the Pediatric Emergency Department at IWK Health in Halifax, Nova Scotia, Canada.

Methods: This was a single-centre, retrospective analysis of administrative ED data from April 1, 2017, to March 31, 2020, from IWK Health Emergency Department in Halifax, Nova Scotia. Triage record data including 101,266 observations of children aged 16 and younger who visited the IWK Emergency Department during a three-year period were used. Several classification machine learning algorithms including Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, and Extreme Gradient Boosting were used to predict patients at high-risk for LWBS. SMOTE was used to handle the class imbalance and improve the performance of the machine learning algorithms. Feature importance was used on the best-performing model to identify the features that are associated with LWBS.

Results: The highest-performing model utilized SMOTE balancing and the XGBoost classification algorithm. Using this model, and data from our partner hospital, an easy-to-follow set of rules were developed for identifying patients at risk of LWBS in real time.

Conclusions: Results show the feasibility of predictive analytics in identifying LWBS patients. This can support proactive decision-making about those patients who are at risk of LWBS.

List of Abbreviations Used

- CTAS: Canadian Triage Acuity Scale
- ED: Emergency Department
- EDWIN: Emergency Department Work Index
- EHR: Electronic Health Record
- FSA: Forward Sortation Area
- ICD10: International Classification of Diseases, Tenth Revision
- KNN: K – Nearest Neighbours
- LWBS: Left Without Being Seen
- ML: Machine Learning
- NACRS: National Ambulatory Care Reporting System
- NEDOCS: National Emergency Department Overcrowding Scale
- RFC: Random Forest Classifier
- RN: Registered Nurse
- ROC AUC: Area Under the Receiver Operating Characteristic Curve
- SHAP: SHapley Additive exPlanations
- SMOTE: Synthetic Minority Oversampling TEchnique
- SVR: Support Vector Machine
- XGBoost: eXtreme Gradient Boosted algorithm

Acknowledgments

I would like to express my gratitude to the following individuals and organizations for their support and assistance throughout my research:

First and foremost, I would like to thank my thesis advisors, Dr. Peter Vanberkel and Dr. Majid Taghavi for their guidance, support, and patience throughout the entire process.

I would also like to thank the members contributing to this work, Dr. Katrina Hurley and Eleanor Fitzpatrick, for their clinical insightfs and feedback, which have helped to shape my research in important ways.

Additionally, I am grateful to the IWK for providing the resources that have enabled me to conduct this research.

Finally, I would like to express my appreciation to my family and friends for their support and encouragement throughout this journey! Especially my family, Diane, Adam and Isabel.

Thank you all for your contributions to this thesis.

Chapter 1: Introduction

Patients who seek care in an Emergency Department (ED) may choose to leave without being seen by a physician. This occurrence is labelled “left without being seen” (LWBS) and is reported in the literature as ranging anywhere from 1.2% to 20.3% [1]. Patients who LWBS may experience an adverse outcome [2]. It has been reported that 11% of adult patients who LWBS require hospitalization within the next week [3]. Recent publications that investigate LWBS have focused on internal hospital factors, patient characteristics, and outcome. Despite the investigation of LWBS occurrences, there are limited studies that research this phenomenon in the pediatric population.

IWK Health (IWK) in Halifax, Nova Scotia, is a women's and children’s hospital that services a catchment area of 1 million people. The IWK Emergency Department (ED) provides care to ill and injured children and youth until their 16th birthday and is classified as an ED Level 1, meaning it covers all aspects of pediatric trauma care, from injury prevention to acute care to rehabilitation [4]. It is the only Level 1 Pediatric Trauma Centre in Canada, east of Quebec. There is motivation from lead clinicians at the IWK ED to gain a more thorough understanding of the LWBS profile present at their institution and to find ways to identify potential LWBS patients in real time.

Most papers in LWBS literature have worked on patient flow metrics such as length of stay, waiting time, and the volume of LWBS. These papers can be separated into three dominant categories:

- External (factors that correspond to patients)
- Internal (factors that relate to procedures or staffing)
- Combination (examining both external and internal factors)

The foundation of external factors LWBS research was completed by Bullard [5], who examined the effects of population growth on LWBS rates over three years. Bullard concluded that while the population grew exponentially in some areas, ED capacity did not incline at the same rate, leading to higher levels of LWBS. It has been shown that LWBS is well correlated to the National Emergency Department Overcrowding Scale (NEDOCS) [6]. Emergency Department Work Index (EDWIN) score has also been investigated for correlation with LWBS frequency, where a strong correlation and discriminatory power for the volume of LWBS in an ED was found [7]. This body of work demonstrates the importance of considering ED load and work when investigating LWBS, as it has repeatedly been shown to be an important factor in LWBS rates.

External factors have also been studied with respect to patient presentation and diagnoses. In a pediatric context, behavioural health presentations in the ED system were associated with higher length of stay (LOS) and LWBS rates. They stated that the presence of behavioural health leads to ED slowdown and increased bed-hold hours, and associated LWBS [8].

Hospital staffing is considered an external factor in LWBS research. For example, a comparison between wait times and LWBS rates between triage systems using nurses and unlicensed assistive personnel concluded that nurses provide higher quality service, lower wait times, and a decreased LWBS frequency [9]. It has also been found that short staffing of nurses led to higher opportunity costs (as represented by LWBS patients) than the cost of increased nurse presence in the ED [10]. In a pediatric context, Gaucher [11] found that the presence of nurse counsellors in the ED led to decreased return visits within 48 hours.

Many papers that focus on a combination of external and internal factors use surveys to identify what would lead a patient to LWBS. McNamara [12] concluded that waiting time was the primary response from respondents when questioned about factors that lead them to leave the ED. It has been concluded that the majority of patients in their study had access to other clinicians elsewhere and highlighted the importance of triage techniques and attention to patient presentation at triage [13]. Table 1 summarizes the literature on LWBS factors.

Table 1. Literature summary for LWBS factor	
Paper	Conclusion
[6]	LWBS correlated with the National Emergency Department Overcrowding Scale (NEDOCS)
[7]	Occupancy rates and emergency department work index scores correlated with LWBS
[8]	Behavioral health census and bed hold hours were significantly associated with increased LOS and LWBS rates
[9]	Wait time decreased and the number of patients who LWBS decreased when nurses performed triage when compared to unlicensed assistive personnel
[8]	Short staffing of nurses led to higher opportunity costs (as represented by LWBS patients) than the cost of increased nurse presence in the ED
[10]	Nurse counsellors in the ED led to decreased return visits within 48 hours.
[11]	Of patients who LWBS, those who receive counseling by a nurse had less return visits in the following 48h.

Table 1. Literature summary for LWBS Factor	
Paper	Conclusion
[12]	Waiting time was the primary response from respondents when questioned about factors that lead them to leave the ED
[13]	Most LWBS had a physician and could obtain care elsewhere

As factors that relate to LWBS occurrences (internal, external, combination) are widely investigated, the prediction of LWBS has not been as thoroughly explored in the literature. One such work is a model by Sheraton [14] that utilizes ED records to develop LWBS profiles with machine learning. The authors used exploratory analysis and model creation to predict LWBS on a dataset that contained over 32 million ED records over one year. Sheraton evaluated the 'main effects with interaction 'model to have a concordance of 63.4% and a discordance of 29.8%. Significant interactions were found with ages below 18 years and chronic conditions and ages above 64 years and chronic conditions. A Random Forest Classifier model was able to predict LWBS with a misclassification rate of 0.013. Primary insurance was the most important predictor, which renders this model less relevant in a Canadian context. Operational variables, such as wait time, duration of stay, or ED load, which are known to be factors associated with LWBS [1, 15-18], were not included in the creation of this model. While the model performs well, it was applied to a small portion (1.25%) of the available data. In another work focused on LWBS prediction, Casey [19] explored adult patients at risk of LWBS using machine learning on three years of data and 217, 250 ED encounters in a US urban setting. Gradient boosting methodology was used on electronic triage records, resulting in a model accuracy of 79% and sensitivity of 89%. As in Sheraton's work [14], a prominent predictor was insurance type/status. Table 2 summarizes the literature on predictive work for LWBS.

Table 2. Literature summary for predicting LWBS		
Paper	Approach / Method	Conclusion
[14]	Over one year, 32,680,232 hospital-based ED visits with 466,047 incidences of leaving without being seen were included. Multivariable logistic regression was used to find significant predictors and their interactions. A random forest algorithm was used to determine the order of importance of factors.	Positive predictors for leaving without being seen were male sex, low acuity, and high annual visits. Negative predictors were Medicare or private insurance, weekend visit, age extremes (<18 years and >65 years), and higher income.
[19]	Three years of data spanning 217,150 encounters for 113,400 patients. XGBoost model achieved an AUC of 0.92 with good calibration, 79% accuracy, 89% sensitivity, and 79% specificity.	The most important features included time of day, acuity, insurance type/status, chief complaint, utilization history, and medications

It is evident that there is limited research on the subject of predictive modelling for LWBS occurrences in EDs. There are few investigations that centre predicting LWBS in Canada. Studies that built predictive models in the United States concluded that primary health insurance was an important feature. As this lacks relevance in Canada, it is possible that the profile of a LWBS patient may contain underreported attributes, where health insurance as a deterrent could result in decreased ED usage by populations that could be represented in a Canadian dataset. Another missing component is pediatric-focused work, where most examples of predictive analysis in EDs related to all ages or adults only.

In reviewing LWBS literature few instances of researchers attempting to predict LWBS, especially for pediatric situations, were found. Site-specific interventions for LWBS reduction, including staffing changes and protocol implementation that were developed for single institutions, were common. For the purposes of this research, the review of literature helped to identify factors which are prominent for LWBS rates at the IWK to proceed to further changes. Appendix A highlights the gaps in the literature that this work targets.

The objective of this work is to gain an understanding of the LWBS patient profile at IWK Health, and find ways to identify potential LWBS patients. This thesis includes two studies in partnership with IWK Health in Halifax, Nova Scotia, Canada to understand the LWBS phenomena in a pediatric setting. Study 1 is a brief paper for a clinical audience with the objective to describe a LWBS cohort and identify key LWBS attributes in a Canadian pediatric context through thorough machine learning analysis. Study 2 expands on Study 1 and includes a descriptive analysis of the data set, presentation of the machine learning models, as well as a guide that has been created for the IWK ED using the machine learning algorithms that were implemented. Both study manuscripts are under review at peer reviewed journals. The manuscripts are presented herein verbatim and contain an introduction, review of relevant literature, methodology, and discussion including limitations.

Before presenting these two studies, Chapter 2 includes an overview of the general methodology, data set, and descriptive analysis results pertaining to the work presented in Study 1 and Study 2. Chapter 3 contains Study 1, and Chapter 4 contains Study 2. Chapter 5 concludes both Study 1 and Study 2 by summarizing notable findings, contributions to current knowledge, and general limitations of the work.

Chapter 2: Methods & Population Characteristics

Machine Learning has become a popular technique in its applications to clinical data [20]. For example, lab results or patient vitals can be translated to numerical values that lend themselves naturally to classification or clustering techniques. Administrative data can provide a rich patient picture, however high frequency of categorical, natural language, and human data entry present higher levels of effort to prepare the data for machine learning modelling.

In general, machine learning can be divided into three main components [21]:

- Data Preparation
- Data Exploration
- Model Creation

Data preparation involves cleaning the data set for the most accurate result. This can include removing duplicate data, imputing or removing data that is missing, or reformatting data to conform to a standard. Data exploration provides points of interest for future modelling, and allows for a deeper understanding of underlying trends that exist in the data. Finally, model creation involves selecting appropriate statistical learning techniques to accomplish the desired objective of the model itself.

In this chapter, the methods for approaching the data preparation, data exploration, and model creation that are used in Study 1 and Study 2 are overviewed.

2.1 Data Preparation

IWK Health provided data for all Emergency Department records between April 1, 2017, and March 31, 2020. In total, there were 101,266 patient records where 5,799 were identified as LWBS. Patient information is collected upon triage in the ED by a Registered Nurse (RN) and registration clerk, and includes information provided by the patient, such as demographic information (name, birthday, address, etc.), automated contextual information (date, time at triage, provider assigned, etc.), and information provided by the RN (CTAS, etc.).

Table 3. Key dataset fields and information

Key Fields	Interpretation	Limitations
Triage Date and Time	Time stamp that records the first entry point to the system. All patients who are recorded in the system are triaged.	Patients who do enter the ED and do not get triaged are not recorded as being present in the system. This could account for some loss of LWBS recording.
Disposition Date and Time	Time stamp recorded when the patient status is updated.	When a patient leaves, they may not alert staff in the ED. In this instance the disposition will be updated when the departure is noticed.
Disposition	Code which encompasses patient visit in the system with their exit status.	Disposition codes are not consistent among years.

Figure 1 shows the evolution of a specific electronic patient record field over time, 'Disposition Code'. This is the field that is indicative of patient status resulting in an LWBS flag. The codes in figure one are described as follows: 62 – Left at his/her own risk post-initial treatment; 63 – Left after triage; 64 – Left after initial assessment; 03 – Patient left at own risk following registration & triage. Further assessment did not occur; 04 – Patient left at own risk following registration, triage & further assessment. Treatment did not occur; 05 – Patient left at own risk following registration, triage, assessment by service provider & initiation. From this Figure it can be seen that various administrative changes have occurred in the timeframe of this project, therefore it can be concluded that extracting certain information (such as vital LWBS) is not straightforward and requires some subject matter expertise.

FIGURE 1. LWBS DISPOSITION CODE BREAKDOWN

To prepare data to be used by machine learning models, several steps including data cleaning, data visualization, and descriptive analysis were completed for the data set. As described in this section, the data was cleaned for inconsistencies and reworked for uniformity. Visualizations and new fields were created that combined information from multiple fields.

Patients with postal codes outside of Nova Scotia were removed. Several features were removed due to redundancy/single value repetition, for example, “Residence Code” (redundant), “Provider Service” (for LWBS patients 94% ‘11001 Registered Nurse’; for All Patients 97%

‘00127 Pediatric Emergency Medicine’), and “Mode of Visit” (99% ‘1 Visit (face-to-face)). To perform modelling on categorical variables, label encoding was used to generate numerical alternatives. Some erroneous data were removed and replaced with an imputation (such as automatically populated timestamps containing the UNIX epoch).

FIGURE 2. FORWARD SORTATION AREA (FSA)'S OF NOVA SCOTIA

An example of this would be default timestamps being replaced with the most common value, such as a 90-minute waiting period for a missing ‘Disposition Date Time’. In all, 2% (2,009) of records were removed before analysis. After these steps, the final data set used for our analysis contains 16 features including 3 engineered features. These features can be seen in Table 4.

The engineered fields present in the final data set are “Time in System”, “Load” and “Driving Time”. “Time in System” represents the best approximation to patient time in the system as is available in this data set. It was derived from the difference between two original features: “Triage Date Time” (as arrival time), and “Disposition Date Time” (as departure time). The limitation to this calculation is that disposition time for LWBS is updated as triage nurses are made aware of an early departure, which may lead to a delay in the disposition time being updated. “Load” is calculated similarly. Load is defined as the number of patients in the system. Using “Triage Date & Time” as a new arrival instance, and “Disposition Date Time” as a departure instance, the number of patients in the ED is updated (increased for each arrival and decreased for a departure) at each discrete instance.

To determine the Driving Time for each patient to get to the hospital, distances were first calculated from the hospital to the patient’s Forward Sortation Area (FSA) which is the first three digits in one’s postal code [22]. To determine the driving duration from each FSA to the IWK, we used the Statistics Canada (StatCan) data set to create the geometry of each FSA using ArcGIS software. The generated polygons representing Nova Scotia’s 77 FSAs are shown in Figure 2. Then, the Latitude/Longitude Centroid of FSA polygons was used for calculating driving distance and duration using Google Maps API with average traffic [23].

Table 4. Features of the Final Dataset

Feature Name	Description
Triage Level	CTAS 1 to CTAS 5
Triage Month	1 to 12
Triage Week	1 to 52
Triage Day	1 to 365
Triage Hour	1 to 24
Triage Minute	1 to 60
Triage Day of Week	1 to 7
Gender	0 - Female; 1 - Male
Age	0 to 19

Table 4. Features of the Final Dataset	
Feature Name	Description
Time in System	Time in system in minutes
Access to Primary HC	1 - Family Physician; 2 - Other; 3 - None; 9 - Unknown/Unavailable
Visit Payor	1 - Provincial or territorial; 2 - WCB, WSIB; 3 - Oth prov or territory; 6 - Other federal government; 7 - Canadian resident, self-pay; 8 - Oth country resident, self-pay
Referred From	1 - Self/Family; 2 - Inpatient Service; 3 - Ambulatory Care Service; 4 - Private Practice; 5 - Drug Dependency Service; 6 - Community Health Service; 7 - Residential Care Facility; 8 - Legal Service; 9 - Educational Agency; 10 - Home Care; 98 - Other; 99 - Unknown
Load	Number of patients in the ED
Driving Time	Driving time from
Main Problem (MRDX)	ICD10 code

2.2 Descriptive Analysis

This section handles the exploratory data analysis performed on the cleaned data set.

In the dataset, around 6% of all patients that visited the IWK ED were classified as LWBS. Hsia et. al. [24] reported that LWBS rates range from 0% to 20.3%, with a median percentage of 2.6%. In the largest sample of its kind, Sheraton et al. [14] reported a LWBS rate of 1.27% from a 2016 sample of 32,680,232 hospital-based ED visits (466,047 LWBS incidences).

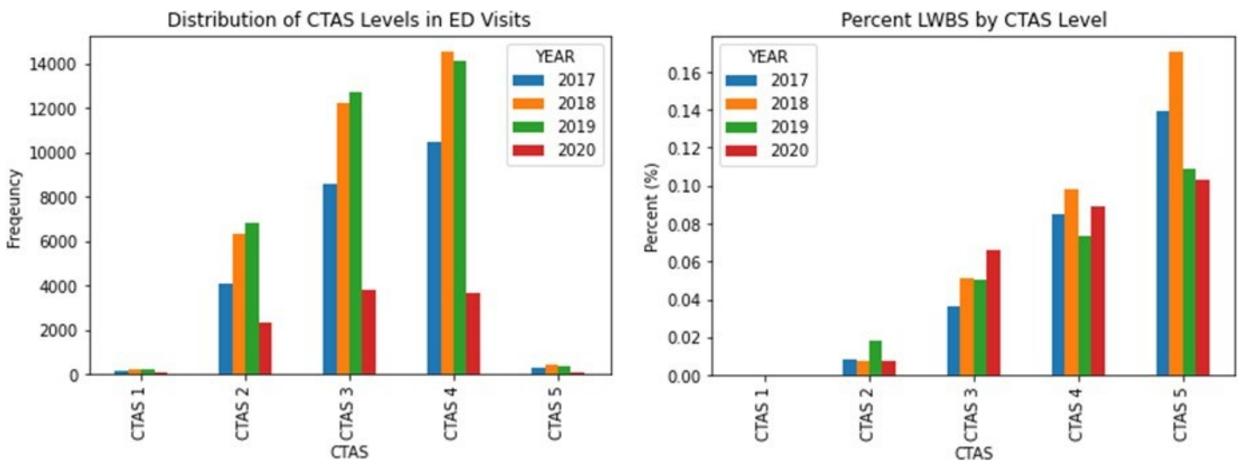

FIGURE 3. DISTRIBUTION OF CTAS LEVELS IN ED VISITS VERSUS LWBS RATE FOR CTAS LEVELS

To explore if patients with higher triage levels are more likely to LWBS, the distribution of the general population is compared to the LWBS population in Figure 3. The IWK ED uses the Canadian Triage Acuity Score (CTAS). CTAS is categorized as 1: Resuscitation, 2: Emergent, 3: Urgent, 4: Less Urgent, and 5: Non-Urgent.

From Figure 3, we see that most LWBS patients are CTAS levels 2, 3, and 4 and as a proportion, there are very few levels 1 and 5. Figure 5 shows the rate of LWBS for each CTAS. The rate was lowest in CTAS 1 and 2 and highest at CTAS 5. The high rate for CTAS 5 shown in Figure 3 (right) and the low proportion of CTAS 5 shown in Figure 3 (left) are caused by the low volume of CTAS 5 patients.

There were no occurrences of CTAS Level 1 patients who LWBS likely due to their high acuity and the need for immediate care. CTAS 3 patients who LWBS are at low risk of harm but are at higher risk to return to the ED [18]. Identifying what types of patients become LWBS, especially those who are more acute is critical as they may experience an adverse event [2]. The literature also demonstrates that CTAS 3 patients who LWBS are at less risk of harm but are at higher risk to return to the ED [25].

The ED characteristics are further explored by the day of the week (Figures 4 and 5) and time of day. Across the dataset, Mondays were the busiest and had the highest rate of LWBS at 7.95% (10,441 visits, 830 LWBS). Saturdays were the least busy and had the lowest rate of

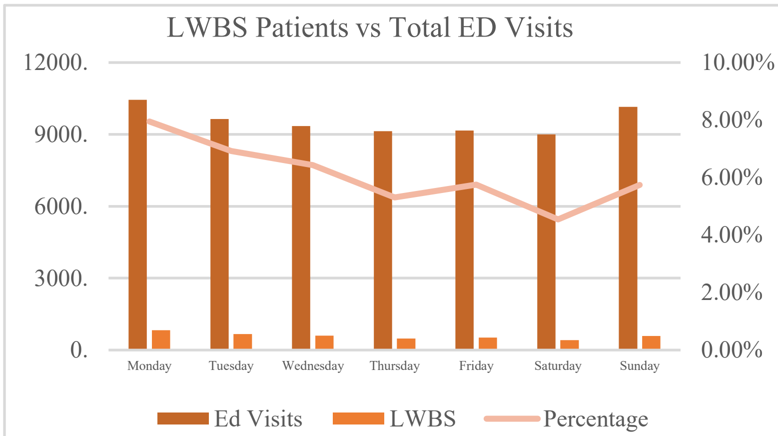

FIGURE 4. LWBS PATIENTS VS TOTAL ED VISITS

LWBS at 4.54% (9,003 visits, 409 LWBS). This is consistent with the literature [14], where patients had lower LWBS instances occurring on weekends.

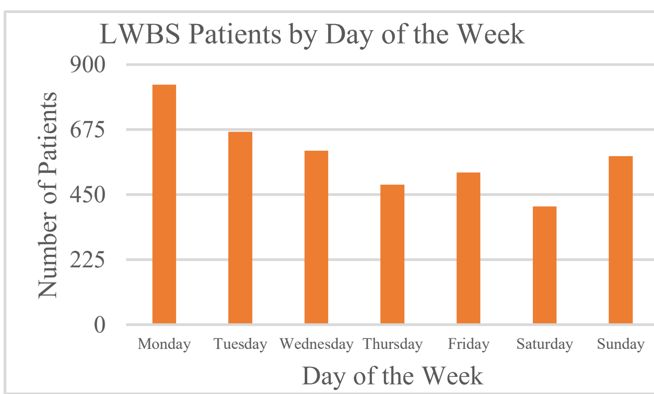

FIGURE 5. LWBS PATIENTS BY DAY OF THE WEEK

Similar to days of the week, time of day was explored (Figures 6 and 7). For this calculation, the triage date and time stamps were treated as entries to the system. It was observed that peak load occurs around 10:00 AM before dropping and then rising again between 4:00 PM and 8:00 PM. LWBS occurrences also increased with ED load but lagged behind.

Crowding and different metrics of load in the ED have been investigated repeatedly in the LWBS literature. Weiss et al. [6] demonstrated that LWBS correlated with the National Emergency Department Overcrowding Scale (NEDOCS), while Kulstad et al. [7] found that the Emergency Department Work Index (EDWIN) score correlated positively with LWBS, and had excellent discriminatory power for the number of patients who LWBS. Vieth and Rhodes [26] found that increased crowding in the ED lead to increased LWBS rates and patient dissatisfaction. IWK ED patients are more likely to LWBS when the load exceeded the average level. In our study, the average load in the ED was 11 patients, and the average load when a patient was identified as LWBS was 19 patients (Figure 8)

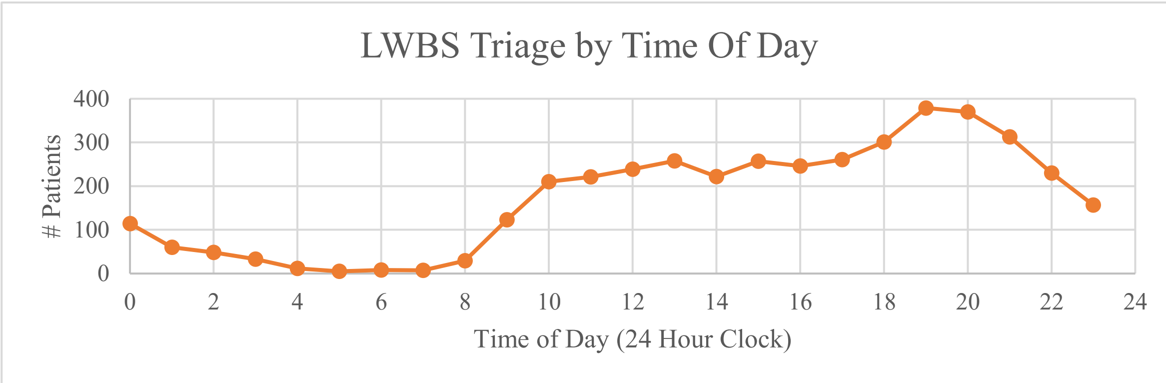

FIGURE 6. LWBS TRIAGE BY TIME OF DAY

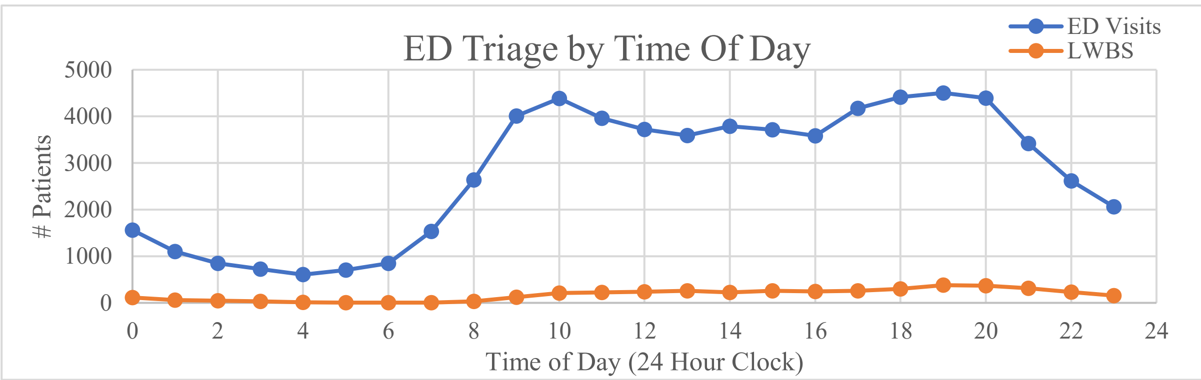

FIGURE 7. ED TRIAGE BY TIME OF DAY

Figure 9 demonstrates the percent LWBS by driving duration, by FSA. Each point on this chart demonstrates a patient from that FSA. There may be other interaction effects that are associated

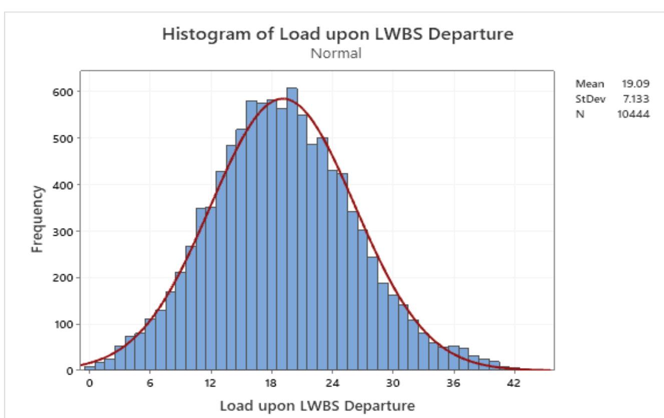

FIGURE 8. LOAD UPON LWBS DEPARTURE

with driving distance, as there is no visually evident trend or distribution. An interesting example of this is above 250 minutes of driving time, where there is a range of 0% to 11% LWBS being observed. There is also a cluster in the 6%-7% range which is reflective of the overall proportion of LWBS in the dataset. It is visually possible to see a slight negative relationship within some of the data, however further analysis will need to be completed to understand the interaction effects of this observation.

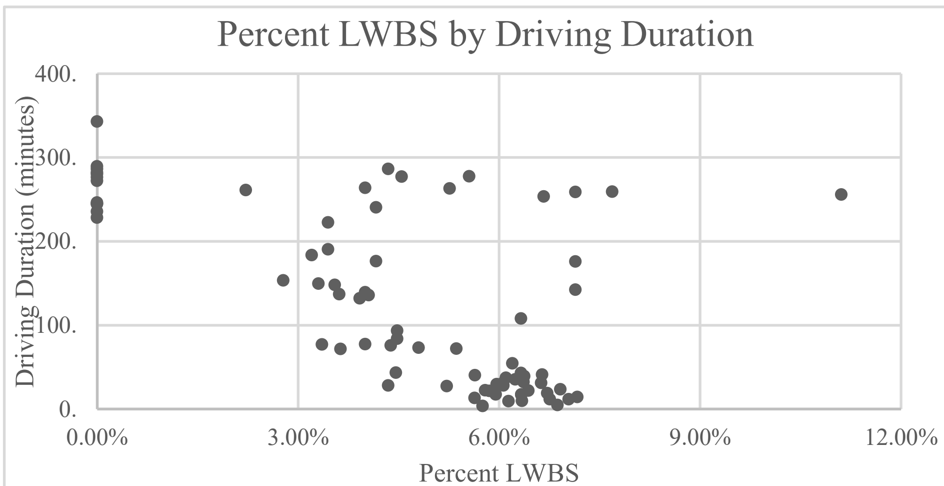

FIGURE 9. PERCENT LWBS BY DRIVING DURATION

2.3 Model Selection

Classification models can help in the determination of what class an observed sample belongs to. In this study, LWBS is the target class that a selected model is aiming to predict. Determining strong features associated with LWBS patients at the IWK ED is also a desired outcome of model creation. The classification models presented in Study 1 and Study 2 are selections of relevant and popular supervised classification techniques. Supervised learning is relevant to this work as there are historical examples of patients being labelled as LWBS that can guide ('supervise') the development of the models for training.

Several supervised classification models were used to predict whether a patient record is likely to be labeled as LWBS or otherwise and compared to determine their efficacy. Numerous classification algorithms have been developed, each with its own strengths and limitations. The choice of algorithm depends on the characteristics of the dataset and the problem being addressed. The model selection stage was refined in three stages across the course of this work. Using the Scikit-learn library for Python, the models that were chosen as candidates were:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- K Nearest Neighbours
- XGBoost Classifier

In some cases, it may be necessary to train multiple models using different algorithms and hyperparameters to choose the best-performing model. This approach is motivated by the fact that different models may perform better on different parts of the dataset, and selecting the best model at the outset of the study is difficult. Model selection is typically performed by evaluating the performance of each model on a holdout dataset, and the model with the best generalization performance is selected.

Logistic regression is a linear classifier that is commonly used when the relationship between the features and the response variable is expected to be linear or close to linear. Decision tree classifier is a simple and interpretable algorithm that is suitable for datasets with easily interpretable relationships between the features and the response variable. These selections are supported by the descriptive analysis where there are strong relationships to LWBS present within certain individual factors (CTAS level, for example). Random forest classifier is a decision-tree-based algorithm that is well-suited for handling non-linear and complex relationships between the features and the response variable. K-nearest neighbor (KNN) is a distance-based algorithm that is suitable for datasets where the relationships between the features and the response variable are not well understood.

XGBoost is an implementation of gradient boosting that has emerged as a popular algorithm for classification tasks. It combines the results of multiple weak models to create a strong model that can make accurate predictions. In XGBoost, each new model is trained on the residual errors of the previous model, with a focus on the samples that were most difficult to predict. This iterative process helps to improve the overall performance of the model, reduce overfitting, and increase the accuracy of the classification task. XGBoost is a powerful algorithm that can handle both numerical and categorical data, and it can be used for both regression and classification tasks. Its speed, scalability, and ability to handle complex datasets with high dimensionality make it a popular choice for data scientists [42].

2.4 Model Performance

Model performance is a crucial aspect of any machine learning task, and it becomes even more critical in cases where the dataset is imbalanced. In such cases, standard metrics such as accuracy might not be sufficient to evaluate the model's performance adequately. The dataset as outlined in Section 2.2 contains only 6% of the target class, which presents a heavy class imbalance. A naïve model that makes a blanket prediction of 0 (non-LWBS) would perform with 94% accuracy. Therefore, it is necessary to use another metric to evaluate performance.

Recall (Equation 1) is a metric that measures the proportion of actual positive samples that the model correctly identified. A high recall score for the minority class implies that the model is effectively identifying instances of this class, even if it means tolerating a higher number of false positives. This is especially relevant in an LWBS scenario, where false negatives represent the highest cost: true LWBS patients that were missed and not identified by the model. Other performance metrics that are calculated in this document are seen in Equations 2 through 4. For feature importance, the F-score was calculated and can be seen in Equation 5. The feature importance plot seen in Chapter 3 (Figure 10) displays the F-score for each feature by considering the number of times a feature is used to split the data across all trees in the model, weighted by the gain in performance (i.e., reduction in the objective function) achieved by each split. The F-scores are then normalized so that they sum to 1, and the resulting values are used to create the feature importance plot (Figure 10).

EQUATION 1 - RECALL EQUATION

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

EQUATION 2 - PRECISION EQUATION

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

EQUATION 3 – ACCURACY EQUATION

$$\boldsymbol{Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}}$$

EQUATION 4 - F SCORE EQUATION

$$F\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

For reliable estimates of model performance, cross validation was used. K-fold cross-validation partitions the data into k equally sized subsets or folds, here k is equal to 10. The model is trained on k-1 folds and tested on the remaining fold. This process is repeated k times, with each fold serving as the test set once. This way, the model's performance can be estimated using the average score across all k test sets, reducing the variance of the performance estimates.

Imbalanced data can require additional augmentation. Synthetic Minority Over-sampling Technique (SMOTE) is a popular method for addressing class imbalance. It creates synthetic samples for the minority class by generating new instances that are similar to the existing minority class samples. The first step is to select the minority class samples that need to be oversampled. Then, SMOTE selects a sample from the minority class and finds its k-nearest neighbors. The synthetic samples are generated by linearly interpolating between the selected sample and its k-nearest neighbors. The number of synthetic samples to be generated can be adjusted by setting the oversampling ratio. After generating the synthetic samples, the resulting dataset can be used to train a machine learning model. This technique has been shown to improve the performance of various classifiers when dealing with imbalanced datasets.

Model refinement was also done using parameter hypertuning. Hyperparameters are parameters that are set before training the model, such as regularization strength, learning rate, and number of trees, which can have a significant impact on the model's performance. This refinement was performed using a grid search for all models. Grid search involves exhaustively searching over all possible combinations of hyperparameters in the search space. Bayesian optimization uses a probabilistic model to iteratively sample hyperparameters that are likely to perform well, based on the previous evaluations of the model. After selecting the optimization technique, the next step is to evaluate the performance of the model for each combination of hyperparameters. This was done using 5 fold cross-validation, where the data is split into training and validation sets and the model is trained and evaluated multiple times with different hyperparameters. Finally, the optimal set of hyperparameters is selected based on the validation performance and used to train a final model on the entire dataset.

For logistic regression, regularization strength (i.e., the penalty parameter) is a key hyperparameter that controls the trade-off between model complexity and overfitting. Similarly, for Random Forest Classifier (RFC) and Decision Tree Classifier (DTC), important hyperparameters include the number of trees, maximum depth, and minimum number of samples required to split an internal node. For K-Nearest Neighbors (KNN), the most important hyperparameter is the number of neighbors used for classification. Finally, for XGBoost, a popular boosting algorithm, hyperparameters such as the learning rate, the number of trees, the maximum depth, and the subsampling rate can have a significant impact on the model's performance. Table 5 shows the first iteration modeling as described in Chapter 4.

Table 5. Hypertuning Results From First Iteration

Model	Parameter	Values	Best
Logistic Regression	C	log(100) to log(0.001)	100
Random Forest Classifier	N Estimators	1 through 1500	100
Decision Tree Classifier	Max Depth	1 through 20	4
K-Nearest Neighbours	N Neighbours	1 through 20	17

All models were trained and tested in an 80/20 train test split. For data that was balanced, original testing sets (non-balanced) were used. Details of the refinement stages, model evaluation, other model tunings, and results are outlined in Chapters 3 and 4.

Chapter 3: Study 1 – Machine Learning to Identify Factors that Predict Patients who Leave Without Being Seen in a Pediatric Emergency Department

Authors: Julia Sarty, Eleanor Fitzpatrick, Majid Taghavi, Peter Vanberkel, Katrina Hurley

Journal: Canadian Journal of Emergency Medicine

Publication Status: Submitted January 18th, 2023.

Contributions:

Conceptualizion: Katrina Hurley, Eleanor Fitzpatrick, Majid Taghavi, Peter Vanberkel

Methodology: Julia Sarty, Majid Taghavi, Peter Vanberkel

Data Acquisition: Katrina Hurley, Eleanor Fitzpatrick, Majid Taghavi, Peter Vanberkel

Data Curation and Formal Analysis: Julia Sarty

Supervision: Peter Vanberkel and Majid Taghavi

Writing – Original Draft Preparation: Julia Sarty

Writing – Review & Editing: All authors

Clinical Insights: Katrina Hurley, Eleanor Fitzpatrick

Abstract

Purpose: To describe a LWBS cohort and identify key LWBS attributes in a Canadian pediatric emergency department through thorough machine learning analysis.

Methods: This was a single-centre, retrospective analysis of administrative ED data from April 1, 2017, to March 31, 2020, from IWK Health Emergency Department in Halifax, Nova Scotia. Variables included: visit disposition; CTAS; triage month, week, day, hour, minute, and day of the week; gender; age; postal code; access to primary care provider; visit payor; referral source; arrival by ambulance; most responsible diagnosis (ICD10); length of stay in minutes; driving distance in minutes; and ED patient load. Descriptive analysis was used to characterise the population. Machine learning modelling was refined iteratively. Dataset balancing and gradient boosting were performed. Model performance was reported using the recall metric.

Results: The dataset included 101266 ED visits where 5800 LWBS (5.7%). The highest performing machine learning model with 16 patient attributes was able to identify LWBS patients with a recall metric of 95%. The most influential features in this model were ED patient load, length of stay (minutes since triage), and driving distance (driving minutes from home address to the ED).

Conclusion: The application of machine learning to administrative ED data successfully produced a model with excellent recall to predict LWBS patients in a Canadian pediatric ED. Further studies are needed to externally validate the model and prospectively evaluate its predictive potential.

Clinician Capsule

Table 6. Clinician Capsule (Study 1)
What is known about this topic?
Up to 17% of patients leave without being seen (LWBS) by a physician in an Emergency Department.
What did this study ask?
Can machine learning modelling create effective models for predicting LWBS using historical data?
What did this study find?
Our machine learning models had excellent recall using data obtained during triage and registration to predict patients that LWBS.
Why does this study matter to clinicians?
This study uses novel methodology and data that would allow it to be applied and evaluated prospectively to predict LWBS.

3.1 Introduction

The mismatch between patient load and Emergency Department (ED) resources for provision of timely care have led to increased numbers of patients who leave the ED without being assessed by a physician (LWBS). Reported proportions of LWBS vary from 1.2% [1] to 16.6% [27]. One

Canadian study found that 2% of children that LWBS experienced an unfavourable outcome [28].

Recent findings with respect to LWBS have found ED crowding was a significant factor, along with younger age, lower acuity and arriving for assessment in the evening and overnight hours [1, 29].

There is little research on predictive modelling for LWBS in pediatric EDs. A machine learning model with over 32 million ED records for patients older than one year of age was developed by Sheraton et al to explore LWBS [14]. The modelling did not account for operational variables such as wait time, length of stay, or crowding. Although it performed well, it was trained on a small proportion (1.25%) of the total available data. Casey [14] used machine learning on three years of adult ED patient data from 217,250 encounters in an urban setting. They used gradient boosting methodology on electronic triage records, resulting in a recall metric of 89%. Both these American models found insurance type and status to be prominent predictors.

The objective of this study was to describe a LWBS cohort and identify key LWBS attributes in a Canadian pediatric context through thorough machine learning analysis.

3.2 Methods

3.2.1 Study Design and Time Period

Single-centre, retrospective analysis of administrative ED data from April 1, 2017 to March 31, 2020.

3.2.2 Study Setting

IWK Health is a tertiary care Pediatric Emergency Department in Halifax, Nova Scotia with an annual census of approximately 35,000.

3.2.3 Dataset

Data were provided in Microsoft Excel by Decision Support Services at IWK Health from the institution's National Ambulatory Care Reporting System (NACRS) dataset. Data were cleaned for inconsistencies and reworked for uniformity. Patients with postal codes outside of Nova Scotia were removed. The following variables were included: visit disposition; CTAS level; triage month, week, day, hour, minute, and day of the week; gender; age; postal code; access to primary care provider; visit payor; referral source; arrival by ambulance; and most responsible diagnosis (ICD10). Three additional variables were derived: length of stay in minutes (difference between triage date and time and disposition date and time), driving distance in minutes (using Forward Sortation Area), and ED load (total number of patients in the ED including those who are waiting to be seen).

3.2.4 Data Analysis and Machine Learning Modelling

Descriptive analysis was used to characterize the population. Machine learning is a type of artificial intelligence that uses historical data to learn and become more accurate at predicting outcomes. Our machine learning modelling was refined iteratively in three stages. First, four relevant classification methods (Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K Nearest Neighbours) were chosen to evaluate base performance. Next, the data set was balanced using Synthetic Minority Oversampling Technique (SMOTE), and the same four models were re-applied. SMOTE was applied due to imbalance of the minority class since only 5.7% of the sample population LWBS. SMOTE generates members of the minority class to augment learning for machine learning modelling. Finally, a powerful gradient-boosted algorithm, XGBoost, was used for the classification of both balanced and unbalanced datasets. These were performed using the Scikit-learn library for Python.

Evaluation for the machine learning models was made using recall to demonstrate how well the model predicted patients classified as LWBS. Recall is a metric that quantifies the number of correct positive predictions made of all positive predictions that could have been made - hence it includes the cases where the disposition was LWBS, and the model predicted otherwise. For imbalanced learning, recall is used to indicate coverage of the minority class [30]. Model features are ranked by F-Score, which represents the number of times a feature has been split on in the model.

3.3 Results

In total, the dataset included 101266 ED visits and 5800 LWBS (5.7%). In all, 2009 (2%) records were removed from the analysis due to missing or inconsistent data.

Lower acuity CTAS patients had higher proportions of LWBS. While there were very few occurrences of CTAS 5 (1093 CTAS 5 visits), these had the highest proportion of LWBS (124, 11.4%). There were no occurrences of CTAS 1 patients who LWBS (597 CTAS 1 visits, 0% LWBS). Across the dataset, Mondays were the busiest day of the week (10441 visits) and had the highest proportion of LWBS (830 visits, 7.95%). Saturdays were the least busy (9003 visits) and had the lowest proportion of LWBS (409 visits, 4.54%). Peak hours for new patient arrivals occurred from 0900-1100 (24399 visits, 24.1% of total visits) and 1700-2000 (17478 visits, 17.3% of total visits). Peak arrival time for patients who subsequently LWBS was 1900-2000 (1049, 18.1% total visits). The age breakdown of LWBS in the ED was as follows: <11 months 678 visits (12%), 1-2 years 1620 visits (0.28%), 3-4 years old (15%), 5-11 years 1705 visits (29%), and >11 years old (18%).

There was no association between driving distance and the likelihood of LWBS on initial visual analysis. For example, patients in areas with greater than 250 minutes of driving had an LWBS range from 0% to 11%.

The mean ED patient load was 11 patients; the average load when a patient LWBS was 19 patients.

The best-performing model from each iteration can be seen in Table 7. The highest recall metric for the best-performing model was 95% for the balanced XGBoost classification model meaning 1 False Negative (patient who LWBS where the model predicted otherwise) to 20 True Positives (the model correctly predicted that the patient LWBS). The feature importance calculated (ranked by F-Score) listed ED patient load as the most influential factor (F-score 558) in the model, followed by triage hour of day (F-score 417), driving minutes (F-score 416), and length of stay (F-score 407).

Table 7. Best-performing models over four iterations (all balanced using SMOTE)	
Model	Recall Metric
Logistic Regression	71%
Decision Tree Classifier	27%
Random Forest Classifier	18%
K Nearest Neighbours Classifier	42%
XGBoost	95%

3.4 Discussion

The highest performing model (recall metric 95%) utilized SMOTE balancing and the XGBoost classification algorithm. While the data used to build this model is drawn from a single ED, the LWBS rates (5.7%) and patterns are typical of those reported in pediatric studies [1, 11, 28]. The size and completeness of our dataset, and the thoroughness of iterative modelling contribute to the strength of this work. Our analysis provides a foundation for larger datasets, particularly those with similar attributes (e.g., pediatric, Canadian), and shows that machine learning could be used to enhance prediction of LWBS with a high recall metric.

ED patient load and waiting time are factors well-understood to be associated with LWBS [6, 7, 31], and load was the most influential factor in our highest performing model. While our model used administrative data, future models could incorporate richer data to capture changes in staffing (clinical and non-clinical), processes of care such as room turnover and physical capacity to examine how these influence LWBS. The capacity to predict outcomes through machine learning would allow EDs to effectively allocate efforts and resources for the greatest impact on patients and families.

Limitations

Limitations relate primarily to the limitations of a historical administrative dataset from a single institution: patients who leave before triage are not captured; departure time for LWBS is often not captured since many patients and families leave without advising a health care provider about their decision to leave the ED affecting both calculated length of stay and ED patient load, and driving time was derived from the patient's home address which may not reflect the address from which they arrived in the ED or the address to which they travelled after departing the ED. Future work will focus on validating the model using broader datasets and prospectively applying this model in conjunction with an intervention to reduce LWBS in our ED.

In conclusion, the application of machine learning to administrative ED data successfully produced a model with excellent recall to predict LWBS patients in a Canadian pediatric ED. Further studies are needed to externally validate the model and prospectively evaluate its predictive potential.

Chapter 4: Study 2 – Using Machine Learning to Predict Patients Who Leave Without Being Seen in a Pediatric Emergency Department

Authors: Julia Sarty; Eleanor A Fitzpatrick; Majid Taghavi; Peter Vanberkel; Katrina Hurley

Journal: International Journal of Medical Informatics.

Publication Status: Submitted on January 30th 2023

Contributions:

Conceptualizion: Katrina Hurley, Eleanor Fitzpatrick, Majid Taghavi, Peter Vanberkel

Methodology: Julia Sarty, Majid Taghavi, Peter Vanberkel

Data Acquisition: Katrina Hurley, Eleanor Fitzpatrick, Majid Taghavi, Peter Vanberkel

Data Curation and Formal Analysis: Julia Sarty

Supervision: Peter Vanberkel and Majid Taghavi

Writing – Original Draft Preperation: Julia Sarty

Writing – Review & Editing: All authors

Clinical Insights: Katrina Hurley, Eleanor Fitzpatrick

Abstract

Background: Patients and their caregivers who seek care in an Emergency Department (ED) may ultimately choose to leave without being seen by a physician. This occurrence is labeled "left without being seen" (LWBS) and can account for up to 15% of all patients who come to an ED. Patients who LWBS do not receive the care they sought in the ED and may experience clinical deterioration related to delayed diagnosis or treatment.

Objective: In this paper, we test machine learning methods to identify which patients are more likely to become LWBS patients. This prediction is intended to be used in practice to prevent adverse outcomes related to LWBS. This paper focuses on the Pediatric Emergency Department at the IWK Health in Halifax, Nova Scotia, Canada.

Methods: We used triage records data including 101,266 observations of children aged 16 and younger who visited the IWK Emergency Department during a three-year period. We utilized several classification machine learning algorithms including Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K-Nearest Neighbors, and Extreme Gradient Boosting to predict high-risk LWBS patients. We used SMOTE to handle the class imbalance in our data set and evaluated the performance of the machine learning algorithms. We used feature importance on the best-performing model to identify the features that are associated with LWBS.

Results: The highest-performing model utilized SMOTE balancing and the XGBoost classification algorithm. Using this model, and data from our partner hospital, an easy-to-follow set of rules are developed for identifying patients at risk of LWBS in real time.

Conclusions: Our results show the feasibility of predictive analytics in identifying LWBS patients. This can support proactive decision-making about those patients who are at risk of LWBS.

4.1 Introduction

An Emergency Department (ED) patient may decide to leave without being seen by a physician or other advanced care provider. Such occurrences are labeled "left without being seen" (LWBS) and often reflect an imbalance between ED resources and patient load. The reported frequency of LWBS can vary significantly in the literature from 1.2% [1] to 16.6% [27]. Patients who LWBS may experience an adverse outcome [2]. It has been reported that 11% of adult patients who LWBS require hospitalization within the next week [3]. A Canadian pediatric study found that 2% of children that LWBS experienced an unfavorable outcome [28]. Recent publications that investigate LWBS have focused on internal hospital factors, patient characteristics, and outcomes. Despite the research on LWBS occurrences, few studies investigate this phenomenon in the pediatric population.

In this study, we partnered with IWK Health in Halifax, Nova Scotia, Canada to understand the LWBS phenomena in a pediatric setting by utilizing administrative data and performing descriptive and predictive analytics on it. There is motivation from lead clinicians at the IWK ED to gain understanding of the LWBS patient profile at their institution and find ways to identify potential LWBS patients before departure. The IWK ED provides care to ill and injured children and youth until their 16th birthday. It is the only Level 1 Pediatric Trauma Centre in Canada, east of Quebec. Level 1 Trauma Centres provide care for all aspects of trauma from injury prevention to acute care to rehabilitation [4].

The rest of this paper is organized as follows. Section 2 introduces relevant literature on healthcare data and predictive modeling, machine learning in ED applications, and the applications of machine learning in studying LWBS. In Section 3, we describe the data set and how it has been cleaned and prepared to be used by the machine learning algorithms. This section is concluded with a descriptive analysis of the data set. In Section 4, we present the machine learning models used to examine the predictability of LWBS and identify the most important features in our data set. Section 5 presents a guide that has been created for the IWK ED using the machine learning algorithms that we implemented. Section 6 concludes the paper with a discussion.

4.2 Literature Review

In this section, we review literature relevant to our study. Section 2.1 reviews some predictive models developed by using healthcare administrative data. Section 2.2 considers machine learning and ED applications, and Section 2.3 focuses on machine learning with application to the LWBS problem.

4.2.1. Administrative healthcare data and predictive modeling

As machine learning intensifies in the healthcare sector, many applications focus on clinical data that results in predictions towards clinical outcomes. Using administrative records can present difficulties as it tends to be categorical and uses natural language, which becomes harder to standardize or clean for future use. However, as administrative data may provide a rounder patient picture, its use is still incorporated in the literature for the prediction of clinical and administrative outcomes.

Gradient-boosted trees are a popular method for developing usable models with administrative health data. This method works by iterating over decision trees and minimizing the loss function. Examples of this method can be seen in predicting postpartum complications [32], predicting 72-hour and 9-day return ED visits [25], prediction of critical care and hospitalization among adults and children [7], prediction of death by suicide within 90 days of an ED visit for para-suicide [33], and by ranking patients by analyzing registration information [34]. Zhang et al. [35] state that gradient-boosted decision trees perform better than other methods such as the traditional least-squares method, ridge regression, Lasso regression, Elastic Net Regression, Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) algorithms.

Machine learning adds utility by improving clinical metrics. Lindberg et al. [36] show that bagging, random forest, and gradient boosting all improve on the Morse Fall Scale for predicting hospitalized patients at risk of falling. Goto et al. [15] surpass the clinical threshold for children during ED triage by incorporating a gradient-boosting method. Using an administrative data set, Desautels et al. [16] uses 8 machine learning algorithms to predict sepsis in an Intensive Care Unit. Sun et al. [17] use routine administrative data to predict the likelihood of hospital admission based on information available at the time of triage. However, Hong et al. [18] found that the addition of historical patient data significantly outperforms administrative data alone in predicting hospital admission from the ED.

4.2.2 Machine learning in ED applications

Predictive and machine learning techniques are being used in the ED set- ting to improve care and increase understanding of current systems, and electronic health records (EHR) are being used to facilitate these investigations. Raita et al. [37] use routinely available triage data to predict ultimate critical care and hospitalization in an eight-year sample set. Using lasso regression, random forest, gradient-boosted decision tree, and deep neural network, the authors found all models to outperform the conventional approach. Goto et al. [15] uses the same predictive methodology in a pediatric subsample, finding that all models developed lead to the higher discriminatory ability of under-triaging critically ill pediatric patients against the conventional clinical approach. These works are limited by their dataset selection and elimination of nearly 10% of records due to incomplete data.

Sanderson et al. [35] uses Canadian ED and patient historical administrative records over a seven-year period to predict death by suicide within 90 days of an ED visit for parasuicide. The resulting gradient-boosted model achieved high discrimination and was an example of thorough ML methodology in an ED context. Hong et al. [25] predicted 72-hour and 9-day return ED visits based on a combination of clinical and administrative data which im- proved on the baseline administrative-only predictions. Hoot [38] created a tool that provides accurate forecasting of output and crowding measures up to 8 hours in the future, which are known factors associated with LWBS. Similarly, Gartner and Padman [39] predicted waiting times in both the waiting room and treatment room and found actual wait time, clinical attributes, and the service environment to be the most important attributes in their model.

There are some examples of using predictive analytics for ED patient flow. Kuo et al. [40] applied stepwise multiple linear regression, artificial neural networks, support vector machines, and Gradient-boosted trees to build real- time ED wait time predictions. This work utilized routine triage data and ED metrics to develop successful models that all outperformed the baseline. Timestamp information was used in feature engineering to understand the flow of patients in the ED.

4.2.3 Machine Learning and LWBS

There have been limited studies that attempt to predict LWBS patients. An exception is a model by Sheraton et al. [14] that utilizes more than 32 million ED records over one year to develop LWBS profiles with machine learning. The authors carried out the investigation using exploratory analysis (development of descriptive statistics for LWBS, bi-variate and multivariate logistic regression to evaluate patient and hospital characteristics against LWBS disposition, and forward step-wise regression to identify significant interactions), and model creation (multiple logistic regression, including main effects with interactions models, and Random Forest Classifier (RFC) on a small subset of observations). Sheraton evaluated the main effects with the interaction model to have a concordance of 63.4% and a discordance of 29.8%. Significant interactions were found with age less than 18 years and chronic conditions and age more than 64 years and chronic conditions. The RFC was able to predict LWBS with a misclassification rate of 0.013. The most important predictor was primary insurance, which is not relevant in a Canadian context. Their model does not account for operational variables, such as wait time, duration of stay, or ED load, which are known to be factors associated with LWBS [2, 5, 6, 9, 7, 12]. While the model performs well, it is applied to a small portion (1.25%) of the available data.

Similarly, Casey et al. [19] explored adult patients at risk of LWBS using machine learning on three years of data and 217,250 ED encounters in an urban US setting. Gradient-boosted trees prediction model was used on electronic health records used at triage, resulting in a model accuracy of 79% and sensitivity of 89%. Like [14], a prominent predictor was insurance type/status. The authors propose a method to integrate model results into the ED work stream, which remains to be prospectively validated. Other relevant machine learning techniques well-suited to administrative and ad- ministrative healthcare data are not attempted in these studies, leaving an opportunity to explore more thorough models for LWBS prediction.

4.2.4 Literature Review Findings

From our review of the literature, we note that machine learning is used in healthcare and ED applications, but little has been done with respect to predictive modeling for LWBS in the ED. The studies noted above lack thorough- ness in machine learning methodology and execution and could benefit from increased model variety as represented in other works surrounding administrative data in healthcare [14, 19]. There is a need to explore these models in a Canadian context, since predominant predictors in existing American studies relate to type and status of health insurance. Finally, most examples of predictive analysis in EDs relate to adults or "all ages" leaving a gap in the application of these models in pediatrics.

This paper works to fill these gaps by applying machine learning techniques to predict LWBS and identify LWBS factors in a Canadian pediatric context. It will also aim to provide relationships and descriptive analysis of the presence of LWBS occurrences at IWK Health in Halifax, Nova Scotia.

4.3 Data Set

The data was provided by Decision Support Services at IWK Health from the institution's National Ambulatory Care Reporting System (NACRS) data set. This includes data for all ED records between April 1, 2017, and March 31, 2020. In total, there were 101,266 patient records where 5,799 were identified as LWBS. Patient information was collected upon triage in the ED by a Registered Nurse (RN) and included information given by the patient, such as their identifiable information (name, birth date, address, etc.), automatic contextual information (date, time at triage, provider assigned, etc.), and information provided by the RN (clinical data). Table 7 demonstrates the baseline characteristics of all patients who LWBS and those who were seen by a physician in the ED. Details on how we prepared the data to be used by the machine learning models as well as a descriptive analysis of the data can be found in Appendix A.

4.4 Machine Learning Models

The two main goals of applying Machine Learning techniques to this data set are to identify the strength of key features associated with LWBS in the ED and to predict whether an individual patient record is likely to be as- signed LWBS. Several supervised classification models were used to predict whether a patient record is likely to be labeled as LWBS or otherwise and compared to determine their efficacy.

The machine learning modeling was refined iteratively in three stages. First, four relevant and popular classification methods (Logistic Regression, Decision Tree Classifier, Random Forest Classifier, and K-Nearest Neighbors) evaluate base performance. Next, the data set was balanced using Synthetic Minority Oversampling Technique (SMOTE) [41], and the same four models were applied. Finally, the Extreme Gradient-Boosting algorithm, XGBoost, was used for the classification of both the balanced and unbalanced datasets. All these methods were performed using the Scikit-learn library for Python.

Evaluation for the base models and all subsequent models was made on the performance of the recall metric, which demonstrates how well the positive (LWBS) class is predicted. The recall metric was chosen as the metric of interest because it assigns the highest cost for False Negatives (true LWBS that the model predicts as remaining in the system). True False Negatives are problematic in our setting and therefore we chose a metric to minimize them. When the data set has an imbalance among classes, recall is used to indicate coverage of the minority class [30].

4.4.1 First Iteration: Unbalanced Data Set

We used the features presented in Section 3 to implement four machine learning algorithms: Logistic Regression, Decision Tree Classifier, Random Forest Classifier, and K-Nearest Neighbors. The data set was randomly split into 80% training samples and 20% testing samples.

All of the machine

TABLE 8. CHARACTERISTICS OF ALL PATIENTS (LWBS OR NOT) IN THE ED FROM APRIL 1ST 2017 AND MARCH 31ST 2020

Characteristic	All Other Patients (%), n = 95468	Patients who LWBS (%), n = 5800
Age Category		
<11 months	14,583 (0.15)	678 (0.12)
1-2 years old	21,942 (0.23)	1,620 (0.28)
3-4 years old	13,243 (0.14)	873 (0.15)
5-11 years old	28,418 (0.30)	1,705 (0.29)
>11 years old	17,282 (0.18)	924 (0.16)
CTAS Level		
CTAS 1	596 (0.01)	0 (0.00)
CTAS 2	19,288 (0.20)	159 (0.03)
CTAS 3	35,493 (0.37)	1,816 (0.31)
CTAS 4	39,148 (0.41)	3,676 (0.63)
CTAS 5	943 (0.01)	149 (0.03)

Table 8. Characteristics Of All Patients (Lwbs Or Not) In The Ed From April 1st 2017 And March 31st 2020

Characteristic	All Other Patients (%), n = 95468	Patients who LWBS (%), n = 5800
Referral by Physician	5,312 (0.06)	249 (0.04)
Patients living near hospital (<5km)	3,179 (0.03)	329 (0.06)
Arrival by ambulance	4,703 (0.05)	85 (0.01)
Season		
Winter	27,115 (0.28)	1,698 (0.29)
Spring	25,353 (0.27)	2,090 (0.36)
Summer	20,745 (0.22)	950 (0.17)
Fall	22,255 (0.23)	1,062 (0.18)
Sex		
Female	44,782 (0.47)	2,799 (0.48)
Male	50,686 (0.53)	3,001 (0.52)
Day of Week		
Weekend	27,425 (0.29)	1,391 (0.24)
Week day	68,043 (0.71)	4,409 (0.76)
Time of Day		
Daytime (8am - 4pm)	48,241 (0.50)	2,519 (0.44)
Evening (4pm - 12am)	37,805 (0.40)	3,026 (0.52)

Table 8. Characteristics Of All Patients (Lwbs Or Not) In The Ed From April 1st 2017 And March 31st 2020

Characteristic	All Other Patients (%), n = 95468	Patients who LWBS (%), n = 5800
Overnight (12am - 8am)	9,422 (0.10)	255 (0.04)

learning algorithms were trained using stratified 10-fold cross-validation to prevent over-fitting and to find the optimal hyper-parameter values. Hyper- parameter tuning was performed using the Grid search method. We created a combination profile for hyper-parameter values for each algorithm. The algorithm's performance was then assessed under each profile using the 10-fold cross-validated area under the curve (AUC), and the one producing the highest result was chosen. Table 9 shows the results from hyper-parameter tuning.

Table 9. Hyper-Parameter Tuning Results

Model Name	Hyperparameter	Value
Logistic Regression	C	100
Decision Tree Classifier	Max depth	4
Random Forest Classifier	N estimators	100
K Nearest Neighbours	N neighbours	17

Table 10 presents various predictive performance measures of the base models. The training and testing scores for all models are high, however, the data set has a minority class imbalance. This means that a naive model (predicting the majority class for every sample) would produce a score of 94%, the proportion of the majority class. This means that despite the high training and testing score, the predictions are not sound. As stated, these models are being evaluated based on the recall metric. The highest recall metric for the base models is 0.214 for the Decision Tree Classifier. This can be interpreted as 10 predicted False Negatives to 3 True Positives. This ratio is not of practical value for our application.

4.4.2 Second Iteration: Balanced Data Set

To improve the results of the first iteration, the same data set used in the previous iteration was balanced using SMOTE and the imbalanced-learn Python library. SMOTE was chosen as a balancing method as it does not duplicate existing rows, rather it imputes data points inside the minority class that are slightly different. This technique uses the K-Nearest Neighbors to augment the minority class and improve prediction quality. The same training and validation methods from the first iteration were used. The results of the second iteration can be seen in Table 11.

Table 10. Performance Metrics For The First Iteration: Unbalanced Data Set

Classifiers	Training Score	Testing Score	True Positive	False Positive	True Negative	False Negative	Precision	Recall
Logistic Regression	0.94315	0.94275	29	27	19040	1131	0.51785	0.025
Decision Tree Classifier	1.00000	0.89790	248	1153	17914	912	0.17701	0.21379
Random Forest Classifier	0.99997	0.94571	111	49	19018	1049	0.69375	0.09690

Table 10. Performance Metrics for the First Iteration: Unbalanced Data Set

Classifiers	Training Score	Testing Score	True Positive	False Positive	True Negative	False Negative	Precision	Recall
K Nearest Neighbours Classifier	0.94712	0.93938	36	102	18965	1124	0.260870	0.031034

Table 11. Performance metrics for the second iteration: balanced data set

Classifiers	Training Score	Testing Score	True Positive	False Positive	True Negative	False Negative	Precision	Recall
Logistic Regression	0.723505	0.706383	828	5607	13460	332	0.128671	0.713793
Decision Tree Classifier	1.000000	0.886093	313	1457	17610	847	0.176836	0.269828
Random Forest Classifier	1.000000	0.939339	207	274	18793	953	0.430353	0.178448
K Nearest Neighbours Classifier	0.926987	0.770801	485	3961	15106	675	0.109087	0.418103

The results of the second iteration show lowered accuracy but far fewer False Negative than the base results displayed in Table 10. After balancing the data and using the balanced training sets, the recall metric was much improved compared to the first iteration across all models. The Logistic Regression model showed a recall metric of 0.714, which can be expressed as this model predicting 2 False Negative for every 5 True Positive.

4.4.3 Third Iteration: Using XGBoost

Even though balancing the data set in the second iteration improved our results, we tried to improve the performance metrics (especially recall) even further. For that purpose, we decided to use the eXtreme Gradient Boosting (XGBoost) method. XGBoost is powerful and reduces processing time in comparison to other models, which could benefit a real-time application for repeated modeling. XGBoost has been a popular and effective method of

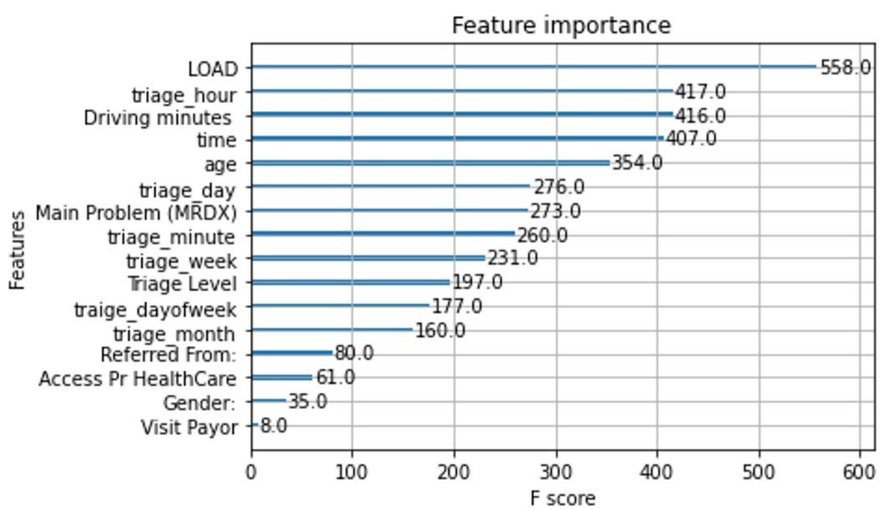

FIGURE 10. FEATURE IMPORTANCE FOR THE XGBOOST MODEL

creating accurate models [42]. XGBoost was applied to the unbalanced and balanced data set along with stratified 10-fold cross-validation. This model produced a recall metric of 0.948, which improved the performance of balanced logistic regression modeling from the previous iteration of modeling (recall of 0.714). This result can be expressed as 20 True Positive predictions for every 1 False Negative. This ratio was acceptable to our clinical partners.

To investigate the functionality of the XGBoost model, feature importance can be pulled to determine what features present in the data set provide the highest prediction value. This provides insight into what the model is weighing as important features in an individual patient being predicted as LWBS.

Figure 10 shows the feature importance of the balanced XGBoost model. The features are ranked by F-Score, which represents the number of times a feature has been split in the model. Load is the metric with the highest F-Score, followed by triage hour, driving distance, and time spent in the ED. From the descriptive analysis, we know that Load is largely correlated with triage hour, explaining why triage hour is elevated as an important feature.

4.5 Applications for IWK ED

The XGBoost model applied to the balanced data set proved to be the best-performing model for our study. The direct application of the method to day-to-day operations was limited as it required ED employees to collect and/or calculate sixteen features for every patient. For these reasons, an easy-to-follow guide that could be used in a case setting is a desirable outcome of the predictive modeling that was performed on the data set.

4.5.1 Guide Creation

To strike a balance between the accuracy of results and implementation effort, we developed a reduced version of the balanced XGBoost model. For this purpose, we only used the most influential features (presented in Figure 10) which were Load (patients currently in the system), driving distance (minutes driving from FSA centroid to the ED), and time in system (minutes from triage to change in disposition status). For completeness and ease of patient segregation, Triage Level (CTAS), was also added to the model. This new model, called the Guide model, utilizes the XGBoost algorithm with SMOTE balanced data, 80/20 data split, and stratified 10-fold cross-validation with the four features defined above. The result of this model was the second-highest performing of all models tested, with a recall of 0.730.

The Guide model was then used to predict LWBS for a range of feature values representing typical patients. An implementable guide is then derived from the results of these predictions. The range of feature values representing typical patients is overviewed in Table 12.

Table 12. Feature spread of sample patients

Feature	Minimum Value	Maximum Value	Step
Triage Level (CTAS)	1	5	1
Time in System (minutes)	0	1,600	10
Driving Time (minutes)	0	340	5
Load (# patients in ED system)	1	47	1

All possible combinations of sample patients were generated from the traits highlighted in Table 12. The table shows the minimum, maximum, and step of each feature. The minimum and maximum represent the boundaries of what was found in the original data set. The step represents how much a feature is increased for each new value. For example, a subset of sample patients with a CTAS value of 5 would contain a patient with all values of the Time in System between 0 and 1600 by an interval of 5 minutes, minutes while holding all else constant. This is repeated for every value of every feature.

The XGBoost Classification model with limited features was executed on these sample patients, along with the assignment of SHapley Additive exPla- nations (SHAP) values for each feature in a sample patient. These SHAP values are model specific and provide insight to feature contributions at all instances of the generated sample population. XGBoost and SHAP values have been used to evaluate feature contribution and model evaluation in relatable contexts [43].

Ultimately, each combination of features yielded a prediction vector (SHAP value) for each feature, a total SHAP value summation across features, and the prediction assigned to the sample patient. To create an interpretable guide, thresholds, where prediction outcomes become positive, were transcribed into a set of rules.

4.5.2 Results

Several thresholds were identified that indicated when a patient was likely to LWBS based on the state of the system (Load) and the patient (time since triage, driving distance, CTAS). The most notable threshold was that it was possible to divide the patient groups into three distinct groups: driving time less than 20 minutes (short), driving time between 20 minutes and 50 minutes (medium), and driving time longer than 50 minutes (long). For each of these groups, the time since initial triage and the corresponding number of patients in the system lead to the guide model predicting an LWBS disposition. It is observed that patients with shorter driving times have a lower tolerance for waiting with increased load. This is externally validated through anecdotal institutional accounts where patients who commit to driving for longer distances are more likely to seek physician consultation. The below steps outline the guide that has been developed:

Step 1: Identify CTAS level.

Step 2: Categorize driving time:

- Short: Less than 20 minutes
- Medium: 20 – 50 minutes
- Long: more than 50 minutes

Step 3: Calculate time since patient triage.

Step 4: Count the number of patients in the system (Load).

Step 5: Use the provided table to determine if patient is likely to LWBS. If the load meets or exceeds the threshold within the time since patient triage, the patient is more likely to LWBS.

Table 13 shows the guide rules for CTAS 3 patients as derived from the prediction model. The data required to apply and interpret these guides are readily available to ED staff allowing the guide to be used directly without information technology support or software development. The guide operates first by identifying what the CTAS score of the patient is (CTAS 1 – CTAS 5). Then, the driving time from the patient's home address - categorized as short (< 20 minutes), medium (20 – 50 minutes), or long (> 50 minutes)- is identified. The final piece of patient information required is how long the patient has been in the system (time since triage). With this information, the appropriate tables can be selected, and the likelihood of LWBS can be identified in conjunction with the current ED load. If the load exceeds the value in the table, the patient in question is likely to LWBS.

4.6 Discussion

The main findings of this work involve the prediction of LWBS likelihood with good recall (how well the positive class is predicted) using thorough machine-learning techniques on administrative data. The highest-performing model utilized SMOTE balancing and the XGBoost classification algorithm with a recall metric of 0.948. This model is institution-specific but provides a foundation for other data sets to be analyzed, especially those with similar attributes (pediatric, Canadian). In addition, demand patterns and rates of LWBS at the IWK ED are typical and representative of the pediatric population. Finally, our analysis showed the predictability of LWBS through machine learning algorithms.

We also expand on the research on predictive modeling for LWBS in EDs. Notably, we approached the modeling from a mathematical perspective expanding on the medical perspective taken by others [14, 19]. An especially practical result from this work is the extension to a relatively

Table 13. Guide table for finding the likelihood of LWBS

Driving Time	Time since patient triage (minutes)	Load (Number of Patients in system)
Short driving time (less than 20 minutes)	<29	5 +
	30 - 39	10 +
	40 -239	16 +

Table 13. Guide table for finding the likelihood of LWBS		
Driving Time	Time since patient triage (minutes)	Load (Number of Patients in system)
Short driving time (less than 20 minutes)	240 - 399	21 +
	400 +	27 +
Medium driving time (20 to 50 minutes)	<29	8 +
	30 - 39	13 +
	40 -239	21 +
	240+	27+
Long driving time (more than 50 minutes)	<29	13 +
	30+	19 +

high-performing guide model using only the most influential features: load (number of patients in the system), time (minutes since triage), CTAS, and driving distance (minutes). The model used to develop this guide achieved the second-highest recall (0.730) with only four features. The guide provides an easy-to-follow set of rules for identifying patients at risk of LWBS.

The models created in this work have face validity and correlate with relevant literature. Load and its variations are hospital factors well-understood to be associated with LWBS [5, 6, 7, 12], and load was the most influential factor both by F-Score in the XGBoost feature selection. The size and completeness of the data set and thoroughness of iterative modeling all contribute to the strength of this work.

Limitations in this work include some assumptions in the engineered features. The data set does not contain any field which would indicate that they are arriving at the ED from a location other than their home address, such as a school or relative's address. Future work in this area may benefit from asking patients from where they are arriving and where they intend to go afterward. Another limitation is the timestamp 'Disposition Time' of LWBS patients since it relies on a triage nurse making note of the patient's departure. For patients who LWBS, this timestamp could be artificially lengthened if the triage nurse is not aware of a departure. For the effect on the features, if disposition time is delayed, the calculation of load would be inflated by one at the time of LWBS departure and a LWBS patient could be recorded as being present longer than they were. However, disposition time was not a field that was observed as consistently an outlier or unpopulated in the data set. Therefore, it is believed that more accurate disposition timestamps would result in further refinement but would not alter the results of subsequent modeling.

In conclusion, this investigation focused on the pediatric Emergency Department at IWK Health in Halifax, Nova Scotia. Using triage records, a guide for ED intervention was proposed after an analysis of descriptive analytics and machine learning methods, including class balancing techniques and feature importance tools. This work can serve as the foundation for further work in pediatric LWBS and predictive analytics for LWBS, including internal validation through the guide model application at IWK Health ED.

Authors' contributions

Research questions were designed by K. Hurley, E. Fitzpatrick, M. Taghavi, and P. VanBerkel. Research design and problem formulation were conducted by J. Sarty, M. Taghavi, and P. VanBerkel. Overseen by M. Taghavi, and P. VanBerkel, J. Sarty conducted the data cleaning and preparation, descriptive and predictive analytics, and also drafted the first version of the manuscript. All authors contributed to the interpretation of results and revised the manuscript. K.

Hurley and E. Fitzpatrick provided critical insights on the clinical and policy implications of the results.

Statement on conflicts of interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Summary Table

TABLE 14. SUMMARY TABLE FOR STUDY 2

What was already known on the topic?
Patients who leave the ED without being seen by a physician are at the risk of clinical deterioration
LWBS patients can account for up to 15% of all ED visits.
Little is known about the effectiveness of machine learning algorithms in predicting LWBS patients at the time of their admission to the ED.
What this study added to our knowledge?
Our study showed the feasibility of using predictive analytics to identify the patients who are at risk of becoming LWBS.
Our study showed that using machine learning, the factors associated with LWBS can be identified.
Knowing the factors associated with LWBS can result in developing guides for hospitals to help them be proactive with LWBS patients.

Chapter 5 : Conclusion

5.1 Discussion

The main findings of this work involve the prediction of LWBS likelihood with good recall (how well the positive class is predicted) using thorough machine-learning techniques on administrative data. The highest-performing model utilized SMOTE balancing and the XGBoost classification algorithm with a recall metric of 0.948. This model parameters are institution-specific but provides a foundation for other datasets to be analyzed, especially those with similar attributes (e.g. pediatric, Canadian). In addition, demand patterns and rates of LWBS rates (5.7%) at the IWK ED are typical and representative of the pediatric population [1, 27, 28]. Finally, our analysis showed the predictability of LWBS through machine learning algorithms.

ED patient load and waiting time are factors known to be associated with LWBS [6, 7, 31], and load was the most influential factor in our highest-performing model. While our model used administrative data, future models could incorporate richer data to capture changes in staffing (clinical and non-clinical), and processes of care such as room turnover and physical capacity to examine how these influence LWBS. The capacity to predict outcomes through machine learning would allow EDs to effectively allocate efforts and resources for the greatest impact on patients and families.

We also expand on the research on predictive modeling for LWBS in EDs. Notably, we approached the modeling from a mathematical perspective expanding on the medical perspective taken by others [14, 19]. An especially practical result from this work is the extension to a relatively high-performing guide model using only the most influential features: load (number of patients in the system), time (minutes since triage), CTAS, and driving distance (minutes). The model used to develop this guide achieved the second-highest recall (0.730) with only four features. These four features represented the four high-importance features by F-Score in the highest performing model. Limiting feature complexity allows for the possibility of real-time guidance that was derived from the highest performing model features. The guide provides an easy-to-follow set of rules for identifying patients at risk of LWBS.

The models created in this work have face validity and correlate with relevant literature. Load and its variations are hospital factors well-understood to be associated with LWBS [5, 6, 7, 12], and load was the most influential factor both by F-Score in the XGBoost feature selection. The size and completeness of the dataset and thoroughness of iterative modeling all contribute to the strength of this work.

5.2 Clinical Implications

The administrative predictive modeling presented in this thesis have important clinical implications for healthcare organizations that wish to leverage data analytics to improve operations and clinical outcomes. In order to build increasingly accurate predictive models, healthcare organizations must have access to high-quality data that is consistent and complete. For the IWK, the implication here is that the move to EHRs is paramount to the furthering of this work. This allows for up to date and maintained models that preserve as much information as possible.

This work utilizes patient triage records to build descriptive and predictive analytics. After using said administrative extensively: several takeaways are relevant for discussion. Firstly, the need for digitize triage records. The current process at the IWK includes a triage nurses recording triage/registration information on paper, before it is collected by a coder and inputted into a spreadsheet format. This results in the ability for human error to interact with the data: forgetting to fill in a field on paper, or a code not being able to interpret handwriting, can easily result in information loss between patient and predictive analyst. Next, there is a desire for fields that are not currently capturesd in the administrative data. These can include but are not limited to: location priod to arrival (other than home address), number of family members present at triage, location after ED (if leaving to seek care elsewhere). A final example is that of patients in the waiting room that are visible to the patient at triage, differering from our definition of load where only patients yet to depart an initial waiting area are counted. These examples represent context at presentation that is lost when dealing with retrospective administrative data, and presents areas for data enrichment that have the potential to significantly imporve administrative data modelling for healthcare.

Administrative predictive modeling can help healthcare organizations improve patient outcomes and optimize resource utilization by identifying patients who are at risk of developing complications or adverse events, such as the LWBS phenomenon in this work. Organizations can proactively intervene and provide targeted preventive care, such as hotspotting in the ED to improve patient outcomes. Hotspotting is the process of identifying high-cost patients in a system. When a patient is predicted as being likely to leave, interventions guided by clinical experts can be performed in an attempt to reduce ED utilization [45]. These could include but are not limited to: checking in with a patient in a waiting room situation, ensuring that they are a good candidate for the ED, or offering wait time estimates.

Resource scheduling can also be improved when more is understood about the dynamics of patient population under investigation, in this case LWBS. Levraging desciptive and predictive analytics for the patient population in the ED, forecasting future demand for ED services can be clarified, allowing healthcare organizations can optimize staffing levels, reduce wait times, and improve patient access to care. Predictive analytics can also be used to identify high-demand periods and allocate resources accordingly, reducing bottlenecks and improving patient flow.

5.3 Limitations

Limitations relate primarily to the limitations of a historical administrative dataset from a single institution; patients who leave before triage are not captured; the exact departure time for LWBS is often not captured since many patients and families leave without advising a health care provider about their decision to leave the ED affecting both calculated length of stay and ED patient load; and driving time was derived from the patient's home address which may not reflect the address from which they arrived in the ED or the address to which they travelled after departing the ED. Future work in this area may benefit from asking patients from where they are arriving and where they intend to go afterward as well as focusing on validating the model using broader datasets and prospectively applying this model in conjunction with an intervention to reduce LWBS in our ED.

5.4 Conclusion

In conclusion, this work showed the feasibility of using predictive analytics to identify patients who are at risk of LWBS. Where there are limited studies that attempt to answer this question, predictive models with high recall were developed using administrative data. Knowing the factors associated with LWBS can result in developing guides for hospitals to help them be proactive with LWBS patients. The guide model developed with four features of strong importance in the highest performing model allowed for the concept of a real-time guide for predicting LWBS patients to be developed.

This investigation focused on the pediatric Emergency Department at IWK Health located in Halifax, Nova Scotia. Using triage records, a guide for ED intervention was proposed after an analysis of descriptive analytics and machine learning methods, including class balancing techniques and feature importance tools. This work can serve as the foundation for further work in pediatric LWBS and predictive analytics for LWBS, including internal validation through the application at IWK Health.

References

[1] J. K. Gorski, T. S. Arnold, H. Usiak, C. D. Showalter, Crowding is the strongest predictor of left without being seen risk in a pediatric emergency department, The American Journal of Emergency Medicine 48 (2021) 73–78.

[2] A. B. Bindman, K. Grumbach, D. Keane, L. Rauch, J. M. Luce, Consequences of queuing for care at a public hospital emergency department, JAMA 266 (1991) 1091–1096.

[3] D. W. Baker, C. D. Stevens, R. H. Brook, Patients who leave a public hospital emergency department without being seen by a physician: causes and consequences, JAMA 266 (1991) 1085–109

[4] Trauma Association of Canada, Trauma system accreditation guideline http://shorturl.at/eFPQS (2011).

[5] M. Bullard, B. Rowe, N. Yiannakoulias, C. Spooner, et al., Recent in- creases in left without being seen in the emergency department, CJEM: Journal of the Canadian Association of Emergency Physicians 4 (2002) 147.

[6] S. J. Weiss, A. A. Ernst, R. Derlet, R. King, A. Bair, T. G. Nick, Rela- tionship between the national ed overcrowding scale and the number of patients who leave without being seen in an academic ed, The American journal of emergency medicine 23 (2005) 288–294.

[7] E. B. Kulstad, K. M. Hart, S. Waghchoure, Occupancy rates and emer- gency department work index scores correlate with leaving without being seen, Western Journal of Emergency Medicine 11 (2010) 324.

[8] Conrad, Heather B et al. "The Impact of Behavioral Health Patients on a Pediatric Emergency Department's Length of Stay and Left Without Being Seen." Pediatric emergency care vol. 34,8 (2018): 584-587. doi:10.1097/PEC.0000000000001565

[9] Paulson, Diane Louise. "A comparison of wait times and patients leaving without being seen when licensed nurses versus unlicensed assistive personnel perform triage." Journal of emergency nursing vol. 30,4 (2004): 307-11. doi:10.1016/j.jen.2004.04.022

[10] Doyle, Stacy L et al. "Outcomes of implementing rapid triage in the pediatric emergency department." Journal of emergency nursing vol. 38,1 (2012): 30-35. doi:10.1016/j.jen.2010.08.013

[11] Gaucher N, Bailey B, Gravel J. For children leaving the emergency department before being seen by a physician, counseling from nurses decreases return visits. Int Emerg Nurs. 2011 Oct;19(4):173-7. doi: 10.1016/j.ienj.2011.03.002. Epub 2011 Aug 23. PMID: 21968409.

[12] K. J. McNamara, Patients leaving the ed without being seen by a physician: is same-day follow-up indicated?, The American journal of emergency medicine 13 (1995) 136–141.

[13] Johnson, Michele et al. “Patients who leave the emergency department without being seen.” Journal of emergency nursing vol. 35,2 (2009): 105-8. doi:10.1016/j.jen.2008.05.006

[14] M. Sheraton, C. Gooch, R. Kashyap, Patients leaving without being seen from the emergency department: A prediction model using machine learning on a nationwide database, Journal of the American College of Emergency Physicians Open 1 (2020) 1684–1690.

[15] T. Goto, C. A. Camargo, M. K. Faridi, R. J. Freishtat, K. Hasegawa, Machine learning–based prediction of clinical outcomes for children during emergency department triage, JAMA network open 2 (2019) e186937–e186937.

[16] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, et al., Pre- diction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach, JMIR medical informatics 4 (2016) e5909.mcn

[17] Y. Sun, B. H. Heng, S. Y. Tay, E. Seow, Predicting hospital admissions at emergency department triage using routine administrative data, Aca- demic Emergency Medicine 18 (2011) 844–850.

[18] W. S. Hong, A. D. Haimovich, R. A. Taylor, Predicting hospital admission at emergency department triage using machine learning, PloS one 13 (2018) e0201016.

[19] P. Casey, K. Zolfaghar, C. Eckert, L. Waters, H. Sonntag, T. McKelvey, N. Mark, 12 predicting patients at risk for leaving without being seen using machine learning, Annals of Emergency Medicine 72 (2018) S5–S6.

[20] R. Bhardwaj, A. R. Nambiar and D. Dutta, "A Study of Machine Learning in Healthcare," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 2017, pp. 236-241, doi: 10.1109/COMPSAC.2017.164.

[21] Miguel A. Hernán, John Hsu & Brian Healy. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks, CHANCE, 32:1 (2019) 42-49, DOI: 10.1080/09332480.2019.1579578

[22] Government of Canada, Forward sortation area—definition, Government Of Canada - Statistics And Research https://www.ic.gc.ca/eic/site/bsf-osb.nsf/eng/br03396.html (2022).

[23] L. McNamara, P. T. Vanberkel, D. Petrie, A. J. Carter, An application and framework for evaluating emergency department networks

[25] W. S. Hong, A. D. Haimovich, R. A. Taylor, Predicting 72-hour and 9-day return to the emergency department using machine learning, JAMIA open 2 (2019) 346–352.

[26] T. L. Vieth, K. V. Rhodes, The effect of crowding on access and quality in an academic ed, The American journal of emergency medicine 24 (2006) 787–794.

[27] N. Gaucher, B. Bailey, J. Gravel, Who are the children leaving the emergency department without being seen by a physician?, Academic Emergency Medicine 18 (2011) 152–157.

[28] Gravel, Jocelyn, Serge Gouin, Benoit Carrière, Nathalie Gaucher, and Benoit Bailey. "Unfavourable Outcome for Children Leaving the Emergency Department without Being Seen by a Physician." Canadian Journal of Emergency Medicine 15.5 (2013): 289-99. Print.

[29] Suastegui, Charles et al. "Comparison of the Demographics and Visit Characteristics of Patients Who Left the Pediatric Emergency Department Without Being Seen With Those Who Were Evaluated in the Emergency Department." Pediatric emergency care vol. 37,6 (2021): e329-e333. doi:10.1097/PEC.0000000000002447

[30] Y. Ma, H. He, Imbalanced learning: foundations, algorithms, and ap- plications (2013).

[31] Stang AS, McCusker J, Ciampi A, Strumpf, Emergency department conditions associated with the number of patients who leave a pediatric emergency department before physician assessment, Pediatr Emerg Care. 2013 Oct;29(10):1082-90.

[32] K. S. Betts, S. Kisely, R. Alati, Predicting common maternal postpartum complications: leveraging health administrative data and machine learning, BJOG: An International Journal of Obstetrics & Gynaecology 126 (2019) 702–709.

[33] M. Sanderson, A. G. Bulloch, J. Wang, K. G. Williams, T. Williamson, S. B. Patten, Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning, EClinicalMedicine 20 (2020) 100281.

[34] L. Luo, J. Li, C. Liu, W. Shen, Using machine-learning methods to sup- port health-care professionals in making admission decisions, The Inter- national journal of health planning and management 34 (2019) e1236– e1246.

[35] Zhang, Bing, et al. Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm. IEEE Access 7. (2019): 32423-32433.

[36] D. S. Lindberg, M. Prosperi, R. I. Bjarnadottir, J. Thomas, M. Crane, Z. Chen, K. Shear, L. M. Solberg, U. A. Snigurska, Y. Wu, et al., Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: a machine-learning approach, International journal of medical informatics 143 (2020) 104272.

[37] Y. Raita, T. Goto, M. K. Faridi, D. F. Brown, C. A. Camargo, K. Hasegawa, Emergency department triage prediction of clinical out- comes using machine learning models, Critical care 23 (2019) 1–13.

[38] N. R. Hoot, L. J. LeBlanc, I. Jones, S. R. Levin, C. Zhou, C. S. Gadd, D. Aronsky, Forecasting emergency department crowding: a prospective, real-time evaluation, Journal of the American Medical Informatics Association 16 (2009) 338–345.

[39] D. Gartner, R. Padman, Machine learning for healthcare behavioural or: Addressing waiting time perceptions in emergency care, Journal of the Operational Research Society 71 (2020) 1087–1101.

[40] Y.-H. Kuo, N. B. Chan, J. M. Leung, H. Meng, A. M.-C. So, K. K. Tsoi, C. A. Graham, An integrated approach of machine learning and systems thinking for waiting time prediction in an emergency department, International journal of medical informatics 139 (2020) 104143.

[41] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.

[42] A. Ogunleye, Q.-G. Wang, Xgboost model for chronic kidney disease diagnosis, IEEE/ACM transactions on computational biology and bioinformatics 17 (2019) 2131–2140.

[43] Y. Meng, N. Yang, Z. Qian, G. Zhang, What makes an online review more helpful: an interpretation framework using xgboost and shap values, Journal of Theoretical and Applied Electronic Commerce Research 16 (2020) 466–490.

[44] Webster, Jane, and Richard T. Watson. "Analyzing the past to prepare for the future: Writing a literature review." MIS quarterly (2002): xiii-xxiii.

[45] Müller-Bloch, Christoph, and Johann Kranz. "A framework for rigorously identifying research gaps in qualitative literature reviews." (2015).

Appendix A

Variation of ‘Concept matrix ’[44] and ‘Framework for Rigorously Identifying Research Gaps in Qualitative Literature Reviews ’[45]

Sources		Concept								Gaps
Author , Year	Name	LW BS	Predic tive Analyt ics	Administ rative Data – sole input	Clinic al Outco mes	Emerge ncy Depart ment	Pedia tric	Hosp ital Proce ss	Thoro ugh model ling	Num ber of Gaps
Casey 2017	Synthetic minority oversamp ling technique for multiclas s imbalanc e problems	X	X	X	X	X			X	2

Sheraton 2020	Patients leaving without being seen from the emergency department: A prediction model using machine learning on a nationwide database	X	X	X	X	X				3
Bourgeois 2008	"Left Without Being Seen": A National Profile of Children Who Leave the Emergency Department Before Evaluation	X		X		X	X	X		3

Kuo 2020	An Integrated Approach of Machine Learning and Systems Thinking for Waiting Time Prediction in an Emergency Department		X	X	X	X			X	3
Conrad 2018	The Impact of Behavioral Health Patients on a Pediatric Emergency Department's Length of Stay and Left Without Being Seen	X		X		X	X	X		3

Sanderson 2020	Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning		X	X		X			X	4

Lindberg 2020	Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: A machine-learning approach		X	X	X				X	4
Raita 2017	Emergency department triage prediction of clinical outcomes using machine learning models		X		X	X			X	4

Polevoi 2005	Factors associated with patients who leave without being seen	X		X		X		X		4
Theiling 2019	Impacting Emergency Department Left Without Being Seen Rates Through Physician Resourcing	X		X		X		X		4

Li 2019	Patients Who Leave the Emergency Department Without Being Seen and Their Follow-Up Behavior: A Retrospective Descriptive Analysis	X			X	X		X		4
Luo 2019	Using machine-learning methods to support health-care professionals in making admission decisions		X	X	X				X	4

Zhang 2019	Health Data Driven on Continuous Blood Pressure Prediction Based on Gradient Boosting Decision Tree Algorithm		X	X	X				X	4
Sun 2011	Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data			X	X	X		X		4

Hoot 2009	Forecasting Emergency Department Crowding: A Prospective, Real-time Evaluation			X	X	X		X		4
Pham 2009	National study of patient, visit, and hospital characteristics associated with leaving an emergency department without being seen: predicting LWBS	X		X		X		X		4

Parekh 2013	Who leaves the emergency department without being seen? A public hospital experience in Georgetown, Guyana	X		X		X		X		4

Weiss 2005	LWBS proportions are used as quality control indicators and this study determined the LWBS proportion at a public hospital in a developing country and some of the patient characteristics associated with LWBS	X		X		X		X		4

Kulstad 2010	Occupancy rates and emergency department work index scores correlate with leaving without being seen	X		X		X		X		4
Paulson 2004	A comparison of wait times and patients leaving without being seen when licensed nurses versus unlicensed assistive personnel perform triage	X		X		X		X		4

Vashi 2019	Applying Lean Principles to Reduce Wait Times in a VA Emergency Department	X		X		X		X		4
McNamara 1995	Patients leaving the ED without being seen by a physician: is same-day follow-up indicated?	X		X		X		X		4
Betts 2016	Predicting common maternal postpartum complications: leveraging health administrative data and machine learning		X	X	X					5

Goto 2018	Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage		X			X	X			5
Al-Stouhi 2015	Transfer learning for class imbalance problems with inadequate data		X	X					X	5
Gartner 2018	Machine learning for healthcare behavioural OR: Addressing waiting time perceptions in emergency care		X			X		X		5

Hong 2019	Predicting 72-hour and 9-day return to the emergency department using machine learning		X			X				6
Hong 2018	Predicting hospital admission at emergency department triage using machine learning		X		X					6
Desautels 2016	Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach		X		X					6

Calster 2018	Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators		X		X					6

TABLE 15. FRAMEWORK FOR RIGOROUSLY IDENTIFYING RESEARCH GAPS IN QUALITATIVE LITERATURE REVIEWS

Category	Sub-category	Definition
Rigorous Modelling		The reason for existence of the research gap
	Model Creation and Selection	There is a lack of appropriate models that were created to address the research question(s)
	Appropriate reference models	Models are not compared to an appropriate reference (i.e. LR, clinical baseline, null classifier)
	Validation technique	There is an absence of internal validity of models/absence of cross-validation technique.
Context		
	Pediatric focus	Papers focus on all ages/adults
Scope		
	Institutional focus	Papers combine many institutions and do not discriminate institutional factors
	Regional focus	Regional factors dominate results from modelling

Appendix B

List of Equations