Machine Learning Predictions of Alternate Level of Care (ALC) in Canada: From Emergency Department to the in-Hospital Stage

Machine Learning Predictions of Alternate Level of Care (ALC) in Canada: From Emergency Department to the in-Hospital Stage

By Faraz Ahmadi, B.Eng.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Science in Computational Science & Engineering

McMaster University MASTER OF SCIENCE IN COMPUTATIONAL SCIENCE AND ENGINEERING (2021) Hamilton, Ontario, Canada

TITLE: Machine Learning Predictions of Alternate Level of Care (ALC) in Canada: From Emergency Department to the in-Hospital Stage

AUTHOR: Faraz Ahmadi, B.Eng.

SUPERVISOR: Dr. Manaf Zargoush, Assistant Professor, DeGroote School of Business

NUMBER OF PAGES: vii, 65

Abstract

In Canada, patients who occupy hospital beds but do not require that intensity of care are called Alternate Level of Care (ALC) patients. ALC has numerous negative implications on patient health and the health care system. Early identification of patients who are at risk of becoming ALC could help decision-makers better manage the situation and alleviate this problem. This thesis evaluates the use of various ML algorithms in predicting ALC at two different time points in the patient's trajectory. Moreover, it identifies the most important predictors of ALC in each time point and provides insights on how adding more information, at the expense of time for decision-making, would improve the predictive accuracy.

Dedication

My sincere thanks go to my thesis supervisor Dr. Manaf Zargoush for his patience and prompt guidance throughout this work. Thanks for believing me and encouraging me in times that I doubted myself.

To my Ph.D. candidate advisor, Somayeh Ghazalbash for her extensive help throughout this project. I would not have been able to finish this work without your guidance.

To my lovely parents, for all their help and support. I am truly grateful for your endless love and encouragement.

To my uncle's family, who supported me from my first moments of arrival to Canada. Thanks for being my home and family throughout these two years.

To all my friends and support system, I am grateful for having each one of you in my life.

Table of Contents

List of Figures

1. Introduction

Delayed discharge has become a prominent challenge in hospitals (Bai et al., 2019). In Canada, hospitalized patients who are occupying beds while not needing the intensity of its services are called alternate level of care (ALC), a term first assigned by the Canadian Institute for Health Information (CIHI) in 1989 (CIHI, 2009). Moreover, a patient is designated ALC only after the physician in charge of their care has decided that the patient is not medically in need of the current acute care setting (Costa et al., 2012; Ahmed, 2019). The majority of these patients are older adults who have received the care and are now waiting to be transferred to a proper facility suitable for their post-acute needs. The "alternate" solutions to staying in acute-care beds could be long-term care (LTC) facilities, rehabilitation centers, or home-care settings (Walker, Morris and Frood, 2009).

Research shows ALC patients who stay in acute care settings have a high chance of developing medical complications such as functional decline, infectious diseases, and depression. These could increase their length of stay (LOS) and make their discharge planning more complicated (Bender and Holyoke, 2018; Arthur et al., 2021).

Delayed discharge affects not only the patient's health but puts a heavy strain on the health system, especially a publicly funded one (Sutherland and Crump, 2013). First, ALC patients occupying acute care beds are costly to the system. They could be staying in more suitable settings with significantly lower costs. In Ontario, it costs an average of $1100/day to stay in acute-care beds, whereas staying in long-term care is less than

$150/day, and in-home care is less than $50/day (Home Care Ontario, 2014). Second, the delayed discharge could act as a bottleneck in the system and adversely affect other health sectors. Hospital beds are finite resources, and when these beds are occupied by ALC patients, it causes the hospital to have less room to treat people in need of those medical services. This would contribute to longer inpatient bed wait times in ED, putting a heavy strain on the ED that is already facing an increase in the number of visits (Milne, Petch and Tepper, 2017). This would, in turn, cause those patients to occupy beds in ED and impede access to emergency beds and other resources for other patients still waiting for care (Health Quality Ontario, 2016).

Moreover, the ALC issue would force hospitals to provide care in unconventional places such as hallways. The rising ALC rate is said to be one of the main reasons for the problem of hallway medicine (Ontario Hospital Association, 2019). Some might even argue that “The loss of beds that results from housing ALC patients is the major cause for congested Emergency Departments, crowded wards, and delayed/canceled procedures and surgeries” (Archer, 2016). Whether ALC is the main reason behind such problems across healthcare sectors or one of the reasons, its impact on patient flow throughout the hospital has been deemed substantial (Costa et al., 2012).

It is estimated that more than 14% of inpatient beds in Ontario are occupied by ALC patients (Bender and Holyoke, 2018). Despite the range of solutions implemented to address ALC by the government and hospitals, the number of patients waiting in hospitals for appropriate care has increased over the years. It is anticipated that the current rise in

ALC rates will continue as Canada is faced with a rapidly growing and aging population (Ontario Hospital Association, 2019). This emphasizes the importance of more assessment around alternate level of care.

There is no one simple solution to ALC (Sutherland and Crump, 2013). It is an interrelated issue among all health sectors, including primary care, home and community care, and acute care hospitals. Most commonly, the ALC patients start their journey by visiting the ED (Bender and Holyoke, 2018). They could be admitted from home, community care centers, or long-term care institutes. After being admitted to the hospital, they receive the care they need before being designated as ALC.

Early identification of ALC patients, beginning at the time of admission to ED, could result in better resource management and decision-making by healthcare managers (Lavergne, 2015). The sooner decision-makers could be informed of patients with complex medical care needs such as those designated as ALC, the better they could start planning their discharge from the hospital.

With the advent of data collection and computing resources, data-driven methods such as machine learning (ML) and statistical modeling could be used to predict ALC designation. However, very few works have been done in this area. More specifically, there has been no research that uses ML to predict delayed discharge at the first patient’s contact point with the healthcare system, i.e., ED. Our aim in this study is to use large administrative datasets gathered throughout the years and machine learning to predict if a patient would be designated ALC in the hospital. Also, linking different data sets from ED and acute care hospitals would result in a data-driven integrated approach to ALC in hospitals. Machine learning allows for simultaneous examination of independent variables in the prediction of ALC. In contrast with statistical methods, ML approaches are more flexible, meaning they could take better advantage of large data sets. Also, ML is more likely to find hidden patterns among predictor variables (Kuhn and Johnson, 2013). It would also help the healthcare planners to make better data-driven decisions.

The research questions posed in this thesis are as follows:

1. Can machine learning be used to predict ALC at different time steps in a patient's journey to the hospital, namely ED and hospital?
2. What are the main predictors of a patient being designated as ALC at each time point?
3. How does adding more information to the model (i.e., from ED to the hospital admission) at the expense of time for decision making improve prediction accuracy? (trade-off between waiting and predictive accuracy)

Early identification of ALC patients long before it happens in acute care would help with optimal discharge planning. Healthcare decision-makers could use these predictions to take more informed actions, such as resource planning, care prioritization, and treatment plans. As the majority of ALC patients are waiting for a long-term care placement (Costa *et al.*, 2012), communications with discharge destinations could be started much earlier to quicken the patient's discharge process. The result could be improved capacity planning both in acute and post-acute care, improved admission /discharge planning, and optimal staffing in healthcare settings. Simply put, information powered by data with enough time to act could alleviate the issue of delayed discharge.

Moreover, finding the predictors of ALC would help identify types of patients who are more at risk of becoming ALC. This could help health care providers and policymakers to take better proactive measures regarding different high-risk populations or patient groups. Also, determining whether a feature in large administrative data sets is associated with ALC or not could enhance future data collections, leading to the collection of more useful predictors. Thus, leading to better prediction capacity.

Finally, predicting ALC designation at two different time points would benefit healthcare providers in a number of ways. First, they can assess the situation at an earlier time point to have enough time for taking necessary actions (e.g., resource planning, etc.). Second, they can wait and use more precise predictions at a later time in the hospital, although it would allow for less time to take action before the patient becomes ALC. Having both options could assist decision-makers in different scenarios.

2. Literature Review

Given the importance of delayed discharge and its adverse effect on the patients' health and health care system, much research has been conducted in this area. In this chapter, the literature around delayed discharge (worldwide) and ALC (in Canada) has been studied. The keywords used to find the related literature were "predict," "delayed discharge," "alternate level of care," "artificial intelligence (AI)," and "machine learning."

In the following paragraphs, we will discuss the related works and categorize them into three main groups, which are (1) identifying determinants of ALC in general or in specific diagnosis groups, (2) deriving clinical risk scores for ALC patients, and (3) predicting ALC as an outcome measure. In the end, we will summarize the conducted research regarding ALC predictions at different time points in the patient's journey in the healthcare system using recent machine learning and show how this thesis aims to fill in the research gaps.

2.1 Identifying determinants of alternate level of care

Earlier studies of alternate level of care, or delayed discharge outside of Canada, have been focused on finding the sociodemographic and clinical characteristics of ALC patients. These studies use conventional statistical methods such as logistic regression to calculate the significance of the factors based on p-value and odds ratio.

Victor et al. (2000) (Victor *et al.*, 2000) studied 456 older adults who were admitted from home to three different elderly care wards in the UK to find the determinants of delayed discharge among them. It argued that factors such as age, sex, and medical conditions were not significant predictors of delayed discharge. In contrast, organizational factors such as care provider type in hospital (e.g., nurse, social worker, or occupational therapist), discharge to institutional care (e.g., nursing homes), and the absence of family caregiver predicted delayed discharge independently.

It must be noted that discharge displacement could not be a predictor of the delayed discharge as the decision that the patient does not require the intensity of current care precedes it. However, Victor et al. (2000) aim to show that organizational factors are more associated with delayed discharge as opposed to patient characteristics. They need to be discharged from an elderly care ward to a nursing home requires a waiting period, resulting in delayed discharge more often.

Challis et al. (2014) considered a similar population of older adults in the United Kingdom, where 665 people were included in this research. They found the discharge disposition type of patient being significantly associated with delayed discharge. This emphasizes once again the effect of organizational factors on delayed discharge. Cognitive impairment was found to be an important clinical risk factor, particularly for those with dementia. Length of acute care stay was also indicative of discharge delay. However, they found that the status of the patient's caregiver before being admitted to the hospital did not affect their delayed discharge. This contradicts the findings of the previous study by (Victor et al., 2000).

More recently, Lenzi et al.(2014) conducted a cross-sectional study over a two-week time span in northern Italy to identify the factors associated with delayed hospital discharge. They assessed more than 6000 patients across a large number of healthcare units with different medical specialties. Patient information consisted of hospital discharge records (age, sex, method of arrival, primary diagnosis, procedures, and comorbidities) and a survey containing living arrangements and dependency metrics (Activities of Daily Living). For patient's diagnoses and diseases, a modified version of Elixhauser disease groups was implemented using ICD-9 codes (Quan *et al.*, 2005). According to this study, an increase in age and number of comorbidities was shown to increase the likelihood of delayed discharge. Moreover, a primary diagnosis of dementia, fracture, or tumor and stay in intensive care units were found to be major factors for delayed discharge.

A series of articles in the literature was centered around studies focusing on the predictors of ALC in certain patient groups. Univariate or multivariate logistic regression is the main analytical tool used in these works. Amy et al. (2012) studied ALC designation among patients with acquired brain injury, a leading cause for death and disability worldwide. Acute care hospitalization records from 2007-2010 in Ontario were used from the discharge abstract data set (DAD) from CIHI. The target variable of having at least one ALC day was analyzed with respect to clinical and demographic variables available in DAD. Multivariate analyses suggested that an increase in age, female sex, having a psychiatric co-morbidity, acute length of stay, and being involved in a motor vehicle collision are among the most significant predictors of ALC designation in acute care hospitals for traumatic and non-traumatic brain injury patients. A finding similar to what was found in other population groups. A key characteristic of this study is that it was conducted across several acute care hospitals; however, because it relied on administrative data regarding the estimates of the ALC days, one limitation of this research is that not all hospitals report ALC days in the same manner and there is a wide variation among ALC days reported by hospitals. It is argued that this leads to underestimation of actual ALC days. Using a relatively large sample size of around 62,000 patients was another strong point of this research.

In another study, Little (2016) investigated the predictors of alternate level of care among mental health inpatient units in Ontario. Logistic regression was applied to features obtained from the RAI-MH data set that is a comprehensive assessment tool for mental health patients. The cut-off point of +30 ALC days was used to define a binary target variable. Little found that dependency (Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) variables), cognitive impairment, social isolation, and being male in addition to older age were among the predictors of delayed discharge from mental health wards.

Stock et al. (2016) also reviewed the population of hypoxic-ischemic brain injury patients who survived acute care in Ontario. They used a multivariable zero-inflated negative binomial regression to identify the predictors of ALC. The authors emphasize the importance of this population as their ALC rates are disproportionally higher than other groups. It was found that 50% of acute care episodes for patients with hypoxic-ischemic brain injury ended with an ALC designation, with a median length of 20 days. In terms of significant predictors, it was concluded that younger age and less severity (measured by time spent in special care units) were unique predictors. A finding that contradicts the previous studies of ALC and the cause for it remained unknown to the authors. Having a previous acute care episode and a psychiatric/behavioral comorbidity were other significant determinants of ALC for these patients. It is noteworthy to say that the sample size of this study was 669 patients.

Turcotte and Hirdes (2015) assessed the long-term (>30 days) ALC stays in Ontario's complex continuing care (CCC) and mental health beds. CCC facilities provide post-acute care rehabilitation and nursing services for patients with complex medical needs. Once again, logistic regression was used to identify the predictors of ALC status. It was found that hospitalization history in the last five years, presence of the spousal caregiver, aggressive behaviors, medical conditions such as Alzheimer's disease and related dementias, and stroke were risk factors of delayed discharge. Functional and cognitive impairment measured by the RAI MDS 2.0 clinical scales was also a significant predictor for ALC stay in CCC beds. The author suggests that future works in this area should focus on the implementation of predictive models to help identify patients with high ALC likelihood from an early stage.

2.2 Deriving clinical risk scores for ALC patients

In the literature, some authors took the work further than just identifying the predictors of delayed discharge and created clinical risk scores for patients in different acute care settings.

Bai et al. (2019) used logistic regression to find predictors of ALC among patients who visited an internal medicine ward in a single hospital in Ontario. After that, they selected the best variables and created a clinical prediction rule, a point system created by assigning a value of 1 to the presence of each predictor in the final LR model and adding them together for each patient. The authors found Age, Sex, Dementia, Diabetes with complications and referral to occupational therapy, physiotherapy and speech-language pathology as their final set of predictors. The derived score was used on the validation set. It consisted of half the total sample size of 4311 patients, achieving an AUC of 0.85. Likelihood ratios for ALC designation were calculated for each of 0 to 6 scores in the derived point system.

In a follow-up study, Turcotte, Daniel and Hirdes (2020) developed a Post-acute Delayed Discharge Risk Scale (PADDRS). An outcome measure designed to identify patients with a risk of delayed discharge at the time of admission to a post-acute care setting. The authors used the previously found predictors of long-term ALC (Turcotte and Hirdes, 2015) and constructed a classification tree to group patients based on the likelihood of delayed discharge. The tree nodes were selected based on the magnitude of the likelihood ratio test statistics and common clinical knowledge. Having an end-stage disease and cognitive performance were the top two classifiers in the decision tree. In the end, patients were clustered into seven distinct risk groups. The data used in this study was partitioned into a train and validation set (70%, 30%), and the validation set of CCC patients across all Ontario achieved an AUC of 0.76. A sample of 30,657 admission episodes from 84 facilities was included in the study. This article was the first study of Alternate level of Care in Canada that tried to predict ALC designation as an outcome measure with patient-level data. The authors discuss that their motivation was to help discharge planners detect patients with a risk of delayed discharge from an early stage. Although the authors acknowledge that their risk score may not have the same discriminatory power at a shorter ALC wait time (less than 30 days), this study shows that predictive models could help decision-makers. Predictive models would not solve the problem of ALC, but having more information and earlier on the patient's journey could certainly alleviate the problem.

2.3 Prediction of ALC status as an outcome

Only one paper was found in the literature to implement a machine learning model to predict delayed discharge. Francis et al. (2015) used neural networks on patients in enhanced recovery following a laparoscopic colorectal cancer surgery to predict delayed discharge and 30-day readmission. The study was conducted in the UK and only had a sample size of 275. The trained neural network achieved an AUC of 0.817 that showed improvement over the use of a multivariable logistic regression with an AUC of 0.807.

In one recent study, Arthur et al. (2021) predicted delayed discharge among home care clients in acute care hospitals. The population of the study is believed to be often frail older adults with complex medical needs. Comprehensive data is gathered from these home care clients using standardized instruments such as the international Resident Assessment Instrument-Home Care (interRAI-HC). The home care reporting system records were linked with the Discharge Abstract Database (DAD) to include information on ALC designation. This cohort data set allowed the authors to have access to many useful measures and predictors that were discussed in the literature but never implemented in a predictive model to find the likelihood of delayed discharge.

Caregiver and marital status, living arrangements, age, sex, medical history and diagnoses in addition to Cognitive scores, instability, pain and frailty scales, Activities of Daily Living (ADL), and Instrumental Activities of Daily Living (IADL) scale from RAI-HC were all used as independent variables in this study. The outcome variable was having 1 or more ALC days in the hospital; this was new in the literature as previous works only investigated longer ALC days (>30 or 7 days)(Francis *et al.*, 2015; Turcotte, Daniel and Hirdes, 2020). Moreover, Arthur et al. (2021) used the largest sample size by far in the literature, a number of 210931 unique home care patients in Ontario and British Columbia.

Their analysis included a descriptive phase which helps select the most significant predictors with respect to the outcome. The select predictors were then deployed in a multivariable logistic regression model. The result was a moderate predictive performance with an AUC of 0.67. The study shows promise that using integrated data sets and information about the at-risk groups prior to hospitalization could be very helpful for care planning and would benefit both the system and the patients. They also mention that the model performance could be improved using machine learning techniques. (Arthur et al., 2021)

2.4 Literature Summary and Identified Gaps

The literature review shows a clear gap in the use of machine learning techniques and large datasets in the prediction of delayed discharge at different time points regarding the patient's trajectory. Although there has been an increase in using machine learning in healthcare settings, such as predicting discharge displacement (Ogink et al., 2019), predicting hospitalization upon entry to ED (Mowbray et al., 2020; Sills, Ozkaynak and Jang, 2021), and predicting admission to Intensive Care Unit from ED (Fernandes *et al.*, 2020), alternate level of care has not been assessed in such ways. These are only a few examples of predictive modeling using machine learning in the healthcare domain. These approaches often outperform statistical methods (e.g., logistic regression).

With little known about the performance and application of ML methods in predicting ALC, we are aiming to find out if ML could accurately predict ALC. For this, we will use retrospective administrative data, linking ED and acute-care hospitals across Ontario to predict ALC designation at different time points in the patient's journey. Our aim is also to take advantage of available large data sets with different ML methods. This study aims to make the following contribution to the literature:

1- Apply a data-driven approach using ML to predict delayed discharge at different time points in a patient's journey, namely in ED before being admitted to the hospital and in acute care hospital

2- We provide insight regarding the trade-off between information and time by adding more predictors with the cost of time for decision making

3- Use retrospective data from 2004-2016 that includes more than 1,400,000 episodes of care. A larger sample size has a better chance of the generalizability of obtained results.

3. Methods

3.1 Study Design

In this study, ML methods were deployed using retrospective electronic health records. Our research is comprised of two separate algorithms at two different time points in the patient's journey. Figure 3.1 presents the overview of the two algorithms. The first algorithm is carried in the ED stage and uses predictors available before admission to acute care hospital to predict delayed discharge in the hospital. The second algorithm uses a combination of predictors in both ED and hospital to predict the same outcome. Emergency department is the first point of contact of patients before becoming ALC in the hospital. Our aim was to use available predictors in ED to predict ALC and then add value to the algorithm by adding more information regarding the patients stay in the acute care hospital. In this chapter, we will elaborate on different steps taken to implement these predictive models in detail. The results are presented in the next chapter.

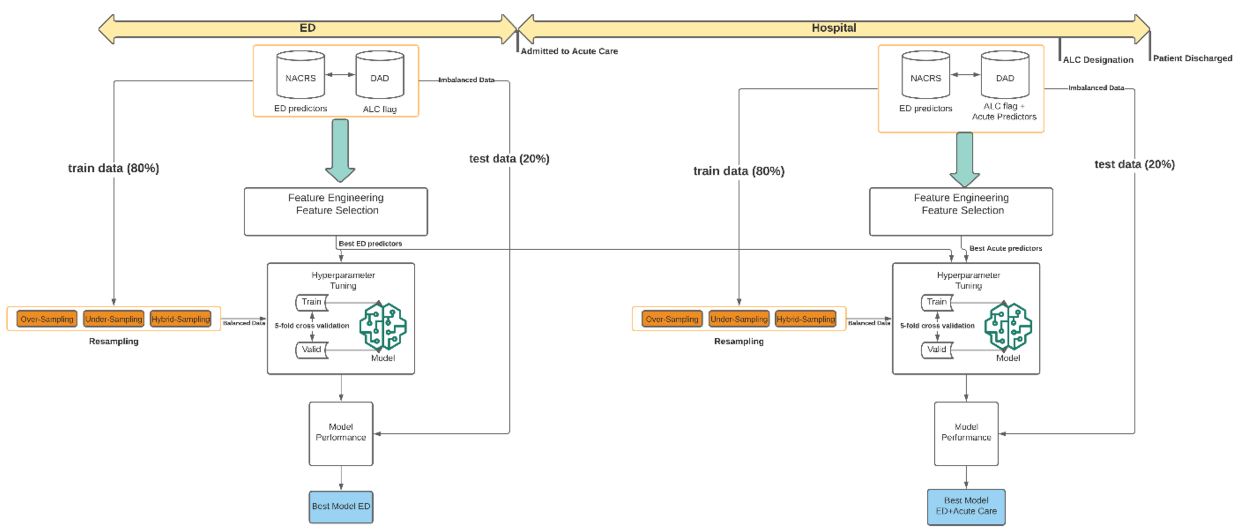


Figure 3.1: The proposed models and their different phases

3.2 Data Sources

We used the National Ambulatory Care Reporting System (NACRS), Discharge Abstract Database (DAD), and Registered Person Database (RPDB). All of these datasets are maintained by the Canadian Institute of Health Information (CIHI). The NACRS contains all visits to facility-based or community-based ambulatory care (CIHI, 2020). This includes visits to ED that are specifically used in this research. NACRS contains all sorts of administrative and clinical variables, marking each patient's journey throughout their visit in ED until being admitted to an acute-care hospital.

The second data set used in this research is a cohort of DAD and RPDP that captures administrative, clinical, and demographic information pertaining to hospital discharges. ALC designation and the number of ALC days in the hospital are also flagged in this data set.

These administrative data sets were provided to us by the Institute for Clinical Evaluation (ICES), accessed remotely via a secure, encrypted VMware virtual desktop server called ICES Data & Analytics Virtual Environment (IDAVE). To access this data, ethical approval was obtained to avoid a data breach and maintain participant confidentiality. Analyses were done solely on this remote system and final outputs were extracted only after being examined for reidentification risk by an ICES analyst.

3.3 Population of Study

For this study, we identified older adults aged 65 and older who visited ED across Ontario and were next admitted to acute-care hospitals between the 2004 and 2016 fiscal years. Therefore, the cohort data set for this study is compromised of the NACRS and DAD data sets, linked together by a unique anonymized ID number provided by ICES. The final cohort consists of 1,474,285 care episodes, of which 174,228 (12%) were designated as ALC patients over their time of stay in acute care.

3.4 Outcome variable

In our predictive models, ALC designation is used as the outcome variable. It is the target variable that the ML predicts based on the features fed to them. From here on, ALC is used as a proxy for delayed discharge from the hospital (Arthur *et al.*, 2021). When a patient is deemed medically fit for discharge by a healthcare provider but stayed in the current setting, awaiting their proper discharge destination, they are assigned an ALC status. The number of days a patient stays in the hospital with this status is coded as the number of ALC days in the DAD data set. In our study, ALC status that was coded as Yes (1 or more ALC days) and No (0 ALC days) is the outcome variable.

3.5 Predictor variables

The NACRS data set provided to us by ICES contained 261 variables. These features captured the entirety of a patient's journey throughout the ED. It contains medical variables such as the mode of arrival, triage score, diagnoses, intervention type, intervention location, provider type, investigative technology (Cat scan, Xrays, etc.) flags, and other administrative variables. It also includes time-based variables of the time of arrival, registration, triage, physician's initial assessment, interventions, and discharge from ED. The demographic variables were present in the cohort data set of DAD and RPDP. The second data set contained additional diagnosis, procedures, length of stay, main service type in hospital, and special care unit information. In the beginning, all independent variables present in the linked data set were considered for the prediction of delayed discharge.

3.6 Data Pre-processing and Preparation

A series of steps were taken to modify and select the predictors of delayed discharge. At first, administrative variables that contained no predictive information (e.g., patient, hospital and health provider unique IDs) were removed from the analysis. The remaining variables were pre-processed, as elaborated in the next paragraphs.

3.6.1 Feature Engineering

The data contained a large number of variables that could not be used with their raw format in ML algorithms. A good example were *dx10code1 – dx10code10* variables that contained diagnosis codes. They were based on the International Classification of Diseases 10th Revision Canadian version (ICD-10-CA) codes. These codes mark a patient's main diagnosed problem and other comorbidities assessed in ED or before. Each variable could contain more than 5000 unique diagnosis codes. Using them with their raw format could be problematic in most ML algorithms, and creating dummy variables for all levels could leave us with the so-called "curse of dimensionality" (Altman and Krzywinski, 2018).

A common approach to resolve this issue in Electronic Health Records (EHR) is to develop risk indices or scores. These include the comorbidity-based Charlson index (Charlson *et al.*, 1987), Elixhauser index (Elixhauser *et al.*, 1998) and their updated versions (Sundararajan *et al.*, 2004). The hospital frailty risk score (HFRS) is another measure that focuses on older adults in acute care settings (Gilbert *et al.*, 2018). However, with the recent advancements in computational power and machine learning systems, more researchers have been trying to grow their medical data feature sets beyond the carefully crafted comorbidity lists (Tran *et al.*, 2014).

Risk scores would not capture all the information available in diagnosis codes and one-hot-encoding (dummy variable) results in sparse and high-dimensional vectors; therefore, a common approach is to leverage the hierarchical arrangement of ICD-10 codes, which means to use only higher-order and less specific of these 5-7 digits long codes. We tried grouping the diagnosis codes by the first three digits and used them in univariate analyses to identify the most associated levels with ALC designation. This enabled us to enhance our models with more useful predictors. Similar approaches to diagnosis codes were taken for other predictors that contained a large number of levels. In some cases, the top levels in terms of occurrence were selected, and the rest were coded as the "other" category.

3.6.2 Feature Selection

Feature selection is the process of reducing the number of predictors used in predictive models. Moreover, "Feature selection is primarily focused on removing non-informative or redundant predictors from the model" (Kuhn and Johnson, 2013). Reducing the number of predictors helps ease the computation cost and also helps with the predictive performance in some cases. It is an important part of every machine learning problem. In supervised classification, there are three most common feature selection methods:

1. Wrapper methods
2. Filter methods
3. Embedded methods

Wrapper methods aim to find the best subset of predictors by creating many predictive models with different combinations of predictors. They use a greedy search method and are computationally intensive. Filter methods use statistical measures outside of predictive models to assess the relationship between each feature and the target variable (Kuhn and Johnson, 2013). The chi-Square test measures the independence of two variables and mutual information uses entropy to determine a score between two random variables. Both are examples of the filter method.

In some algorithms, the feature selection process is embedded automatically in the model training. "Some models are naturally resistant to non-informative predictors. Tree- and rule-based models, MARS and the lasso, for example, intrinsically conduct feature selection" (Kuhn and Johnson, 2013). These are called embedded or intrinsic feature selection methods.

A comprehensive literature review and the statistical significance from the univariate analyses were combined in our approach to identifying the predictors of ALC. In the literature, there was no clear link between the ED visits and ALC; therefore, in the beginning, we tried to produce as many meaningful features as possible that were indicative of the patient's process through ED. Then, a systematic approach was taken to select the best variables. First, each variable was individually investigated. The frequency and missingness of each were reported. Variables with little or no variance between their levels and a high amount of missingness were excluded from the analysis. Variables not pertained to our outcome measure or those which were recorded after the ALC incidence in acute care were also removed.

Subsequently, the remaining variables and engineered features were implemented in univariate analysis with respect to ALC. Then the most significant features were selected. The goal was to include as many features as possible due to the absence of a clear link in the literature between ED variables and ALC. Finally, the mutual information (MI) criterion was used to rank the selected features in both ED predictors and after adding hospital predictors. ML algorithms were implemented using different subsets of ranked features (by MI) to identify the optimal number of predictor features. We tried not to include too many features as it complicates the algorithm while attempting to achieve higher accuracy.

3.6.3 Imbalanced Data Resampling

In practice, a common problem in supervised classification problems is that we have to deal with imbalanced data sets, i.e., the data is not evenly distributed between target classes and that there is a majority (referred to as "negative" in medical cases) and a minority (referred to as "positive") class. For imbalanced classification, the interest usually leans towards the correct classification of the rare positive class. This poses a challenge for predictive modeling as most of the ML algorithms used for classification were designed around the assumption of an equal number of examples for each class. Therefore, classic metrics such as total error or accuracy are not useful in imbalanced data sets. For example, in a 2-class data set when the majority takes up 95% percent of the data, the ML model could easily assign all observations to this class and achieve a high accuracy of 0.95, whereas for the 5% sample that was important it is achieving a per-class-error-rate of 100%.

There are two common approaches to eliminate the problem within imbalanced data sets. One is based on adjusting the cost function in ML models. Assigning a high cost to misclassification of the minority class while trying to minimize the overall cost (Krawczyk, Woźniak and Schaefer, 2014; Domingos, 1999). The other approach modifies the training sample fed to ML models. It uses resampling methods, either over-sampling the minority class or under-sampling the majority class, or a combination of both (hybrid-sampling). In addition, the synthetic minority over-sampling technique (SMOTE) is sometimes used to create synthetic instances of the minority (positive) class to balance the imbalanced data (Fernandes *et al.*, 2020; Chawla *et al.*, 2002).

In our cohort data, ALC patients consist of 12% of data, leading to an imbalanced classification problem. Up-sampling, down-sampling, and SMOTE techniques were used on the training data before feeding them to each model. The test data remained imbalanced. SMOTE was not used in the final models as it did not improve the accuracy. In almost all cases, over-sampling the training data performed better than other resampling techniques.

3.7 Analysis

We used R version 3.6.3 for all analyses. Predictive models were mostly built on *H2O*, which provides a faster, more efficient tool for building ML models on large data sets. *H2O* uses in-memory processing with fast serialization between nodes and clusters to support massive datasets. Moreover, its distributed processing on big data delivers speeds up to 100x faster with fine-grain parallelism, enabling an overall optimal solution for building ML models in R (*R Interface for H2O*, 2020). The *caret* package was also used for implementing some ML models that were not available in H2O (Kuhn, 2008).

After engineering and selecting the final predictors for both models, stratified random sampling was used to split the data set into the train (80%) and test (20%) sets. The data set was pre-processed before this stage; however, the test set was remained imbalanced to be indicative of true ALC distribution. In the training set, a stratified 5-fold cross-validation was performed to find the best hyperparameters. The area under curve (AUC) was the predictive accuracy criteria for hyperparameter tuning. It is important in ML that the test data is unseen by the model until it is fully tuned with the best hyperparameters. Cross-validation is used to avoid overfitting and ensure better generalizability during the tuning phase.

We used Logistic Regression (LR), Elastic Net (EL), Classification and Regression Tree (CART), an ensemble of decision trees with bagging (treebag), random forest (RF) and gradient boosted trees (GBT). We aimed to use a variety of tree-based and non-tree-based ML models.

3.8 Performance measures

When dealing with imbalanced data sets, overall accuracy is not a good indicator of the model's performance; therefore, we used 7 different classification metrics. We used sensitivity (recall), precision, specificity, and F1-score. All these metrics are a function of the confusion matrix shown in Table 3.1. Sensitivity measures the proportion of those who were designated ALC (positive class) that were classified as ALC by the algorithm. Specificity refers to the proportion of those who were not ALC that were not classified as ALC (negative class) by the algorithm. We also used the area under the receiver-operating curve (AUC), a method commonly reported in the literature (Francis *et al.*, 2015; Bai *et al.*, 2019; Arthur *et al.*, 2021). ROC is a function of true positive ratio (TPR) or recall versus the false positive ratio (FPR) plotted over all possible classification thresholds. An AUROC of 0.5 is achieved with a trivial random classifier, whereas 1 represents perfect discrimination among classes.

Another great measure of success for imbalanced classification is the Area Under Precision-Recall Curve (AUPRC). The AUPRC shows the trade-off between precision and recall among different thresholds. Higher values indicate better performance in classifying the minority class. It must be noted that AUPRC could not be used to compare models trained on different data sets with different imbalanced ratios among classes. But AUPRC is a good measure in comparing different models' performance on one data set.

F1-score was also assessed, and it is a great measure used in evaluating classification models suited for dealing with imbalanced data sets. It is defined as the harmonic mean of precision and recall, as shown in Table 3.1. In *H2O*, F1-score is automatically used to select the best threshold.

Every classification model assigns a probability (from 0.0-0.1) to each class; it is our job to come up with the best threshold used to assign class labels. The default threshold for interpreting probabilities to class labels is 0.5; however, in imbalanced data, it is not the best decision. In order to find the optimal threshold, we have to decide on our desired metric or metrics and seek a threshold that maximizes that metric. In this study, models were trained and evaluated using AUC and AUPRC during the hyperparameter tuning stage. Also, the best threshold was selected as the one achieving the highest F1 score among the rest.

Table 3.1: Confusion Matrix and common performance measures

	Predicted Negative Class	Predicted Positive Class
Actual Negative Class	True Negative (TN)	False Positive (FP)
Actual Positive Class	False Negative (FN)	True Positive (TP)

$$True\ Positive\ Rate\ (TPR), Recall, Sensitivity = \frac{TP}{TP\ +\ FN}$$

$$False\ Positive\ Rate(FPR) = \frac{FP}{FP\ +\ TN}$$

$$Precision = \frac{TP}{TP\ +\ FP}$$

$$Specificity = \frac{TN}{TN\ +\ FP}$$

$$F1 - Score = \frac{2\ \times\ Precision\ \times Recall}{Precision\ + Recall}$$

3.9 Hyperparameter Tuning

Machine learning algorithms require parameter tuning to ensure maximum performance. Each algorithm has its own set of parameters, and the optimal parameters could significantly affect the resulting model's performance (Claesen and De Moor, 2015). The following table shows the hyperparameters each algorithm had tuned in this study.

Table 3.2: Description of Gradient Boosted Trees Hyperparameters

Gradient Boosted Trees	
Max depth	Maximum tree depth
Sample Rate	For each tree in the ensemble, we only use a sample of the balanced training data.
Columns sample rate	The rate at which the sampling of training data's features (columns) occurs at every split node.
Column rate per tree sample	Another sampling of columns at each tree. It is multiplicative with *column sample rate,* meaning if both are set to 0.8, it results in 64% of columns being used at each split.
Column sample rate change per level	Specifies the change in column sampling as a function of tree depth.
Min rows	Indicates the minimum number of observations needed at a leaf to split. For example, if min_rows =512, at each node, we need more than 512 responses on both Yes and No values of the outcome variable (ALC status).
Min split improvement	The minimum relative decrease in error needed for a split to happen at any given node. If a split does not improve the error rate by the specified amount, it will not happen.
Histogram type	The histogram aggregation method used to find the best split point.
Learning Rate	The rate at which GBT learns when building model
Learn rate annealing	Reducing the learning rate after each tree by this factor
Early stopping criteria and metric	The criteria and metrics (AUC, AUPRC, F1 etc.) used for early stopping in ensemble methods.

Table 3.2 shows in detail the hyperparameters in gradient boosting trees. In GBT , the training data could be sampled on both rows (observations) and columns (features), causing the trees in GBT to not use all the training data. This might reduce the training performance but often helps improve the test accuracy (Friedman, 2002). The early

stopping criteria used was based on AUPRC. Meaning if the AUPRC did not improve after 10 consecutive scoring events (each scoring event is 10 tree intervals apart), the model would not build any more trees and stop the process there. Table 3.2-Table 3.5 show the hyperparameters for Random Forest, Elastic Net, and CART. Logistic Regression and Bagging CART were used but had no hyperparameters.

Table 3.3: Description of Random Forest Hyperparameters

Random Forest	
Max depth	Maximum tree depth
Mtries	Number of splitting variables randomly selected at each level
Min rows	Indicates the minimum number of observations needed at a leaf to split. For example, if min_rows =512, at each node we need more than 512 responses on both Yes and No values of the outcome variable (ALC status)
Min split improvement	The minimum relative decrease in error needed for a split to happen at any given node. If a split does not improve the error rate by the specified amount, it will not happen
Sample rate per class	For each tree in the ensemble, we only use a sample of the balanced training data. This sample rate is different for members of each class in the training data
Histogram type	The histogram aggregation method used to find the best split point
Early stopping criteria and metric	The criteria and metric (AUC, AUPRC, F1 etc.) used for early stopping in ensemble methods

Table 3.4: Description of Elastic Net Hyperparameters

Elastic Net	
Lambda	Controls the penalty strength for both regularizations. If lambda = 0, no regularization is applied and alpha is ignored.

Alpha	The alpha parameter controls the distribution between the $\ell 1$ (LASSO) and $\ell 2$ (ridge regression) penalties.

Table 3.5: Description of CART Hyperparameters

Classification and Regression Tree (CART)	
Tree depth	Depth of the implemented tree. Also, a measure of the number of predictors involved in the tree as tree nodes.

4. Results

The results from the predictive models are presented in this chapter.

4.1 Descriptive Analysis

The results for the descriptive analysis on patients who visited the ED and were admitted to the hospital between 2004 to 2016 are depicted in Appendix A: Descriptive Analysis Table. The data contained a total of 1,474,285 episodes of care and 765,133 unique patient IDs. The mean age of our sample was 75 years (SD 7.8). Out of this population, 174,228 (12%) episodes of care resulted in ALC designation compared with 1,300,057 (88%) episodes of non-ALC designation. The mean age for ALC designated population was 78 years (SD 8.0), whereas it was 75 years (SD 7.7) for non-ALC visits. The majority of ALC patients were females (58%). Most of them stayed in urban hospitals (87%) and four out of every five ALC patients (80%) entered ED via ambulance. Approximately 93% of all visits by ALC patients were deemed to be urgent, receiving a Canadian Acuity and Triage Scale (CTAS) score of three or less. Missing data percentage was less than about 5% for all variables included.

4.2 Hyperparameter Tuning results

Table 4.1 summarizes the results of hyperparameter tuning for the CART algorithm for both models. Table 4.2 shows the range and optimal points for Random Forest hyperparameters. For Gradient Boosting Machine, the learning rate was 0.05, and *learn_rate_annealing* was set to 0.99. This combination ensures convergence as it is both small from the start and shrinks after each iteration. Table 4.3 summarizes the rest of GBT tuned hyperparameters. GBT and RF were deployed using *H2O*. Encoding of categorical variables is an important part of our predictive models as the majority of predictors are categorical. *H2O* uses "Enum" encoding that does not change the data set and internally maps the strings to integers. Then it uses those integers to make splits in trees. The Enum method proved to be more practical than one-hot-encoding for our categorical features. Table 4.4 shows the optimal values for the Elastic Net models.

Table 4.1: Tuned hyperparameters for CART

Models	***CART***	
	Parameters	Tree length
	Range	1-30
ED only	Optimal	4
ED + Hospital	Optimal	5

Table 4.2: Tuned hyperparameters for Random Forest

	Random Forest					
Models	Parameters	mtries	Min rows	Max depth	Min split improvement	Histogram type

	Range	1-36	2^x: $x \epsilon \{0\ to\ (\log_2 N^{i}) - 1\ by\ 1\}$	5 - 30	0, 1e-8, 1e-6, 1e-4	Uniform Adaptive, Quantiles Global, Round Robin
ED Only	Optimal	10	64	17	1e-04	Quantiles global
ED + Hospital	Optimal	16	64	18	0	Quantiles global

Table 4.3: Tuned hyperparameters for GBT

	Gradient Boosting Trees					
Models	Parameters	Sample rate	Min rows	Max depth	Min split improvement	Column sample rate
	Range	0.4-1 by 0.01	2^x: $x \epsilon \{0\ to\ (\log_2 N^{ii}) - 1\ by\ 1\}$	4-20	{0, 1e-8, 1e-6, 1e-4}	0.4-1 by 0.01
ED Only	Optimal	0.92	512	14	1E-06	0.45
ED + Hospital	Optimal	0.79	256	12	0	0.99

	Gradient Boosting Trees - Continued			
Models	Parameters	Column sample rate per tree	Column sample rate change per level	Histogram type

[i] N is training sample size
[ii] N is training sample size

	Range	0.4 - 1 by 0.1	0.9 - 1.1 by 0.01	Uniform Adaptive, Quantiles Global, Round Robin
ED Only	Optimal	0.9	1.05	Quantiles Global
ED + Hospital	Optimal	0.7	1.09	Round Robin

Table 4.4: Tuned hyperparameters for Elastic Net

Models	***Elastic Net***		
	Parameters	Lambda	Alpha
	Range	{1, 0.5, 0.1, 0.01, 1e-3, 1e-4, 1e-5, 0}	0-1 by 0.01
ED only	Optimal	1e-5	0.39
ED + Hospital	Optimal	1e-5	1

4.3 Predictive Model Performance

Table 4.5 summarizes the various performance measures for each ML algorithm in both ED and acute care models. **Error! Reference source not found.** and Figure 4.2 illustrate the 95% confidence interval of all performance measures across ED and Hospital models. It is based on the 5-fold cross-validated models trained on the optimal hyperparameters. Overall, models based on ED predictors only achieved an AUC of 0.73. After adding more information to the first model and by including more predictors from the hospital, an AUC of 0.81 was achieved. GBT was the best classifier for both ED and Hospital models.

Table 4.5: Performance measure on test data

Model	ML Algorithm	AUC	AUPRC	Recall	Precision	Specificity	F1 Score	Accuracy
Model 1 (at ED)	CART	0.66	0.21	**0.61**	0.19	0.63	0.28	0.63
	Bagging-CART	0.69	0.23	0.52	0.22	0.74	0.31	0.72
	Logistic Regression	0.73	0.27	0.47	0.27	**0.92**	0.34	0.79
	Elastic Net	0.73	0.27	0.45	0.28	**0.92**	0.34	**0.80**
	Random Forest	**0.74**	0.28	0.47	0.28	**0.92**	**0.35**	0.79
	Gradient Boosted Trees	**0.74**	**0.29**	0.46	**0.29**	**0.92**	**0.35**	**0.80**
Model 2 (in hospital)	CART	0.72	0.25	0.49	0.27	0.82	0.35	0.78
	Bagging-CART	0.76	0.27	**0.61**	0.26	0.77	0.37	0.75
	Logistic Regression	0.79	0.33	0.55	0.32	0.93	0.40	**0.81**
	Elastic Net	0.79	0.33	0.55	0.32	0.93	0.40	**0.81**
	Random Forest	**0.81**	0.36	0.59	0.33	**0.94**	0.42	**0.81**
	Gradient Boosted Trees	**0.81**	**0.37**	0.57	**0.34**	**0.94**	**0.43**	**0.81**

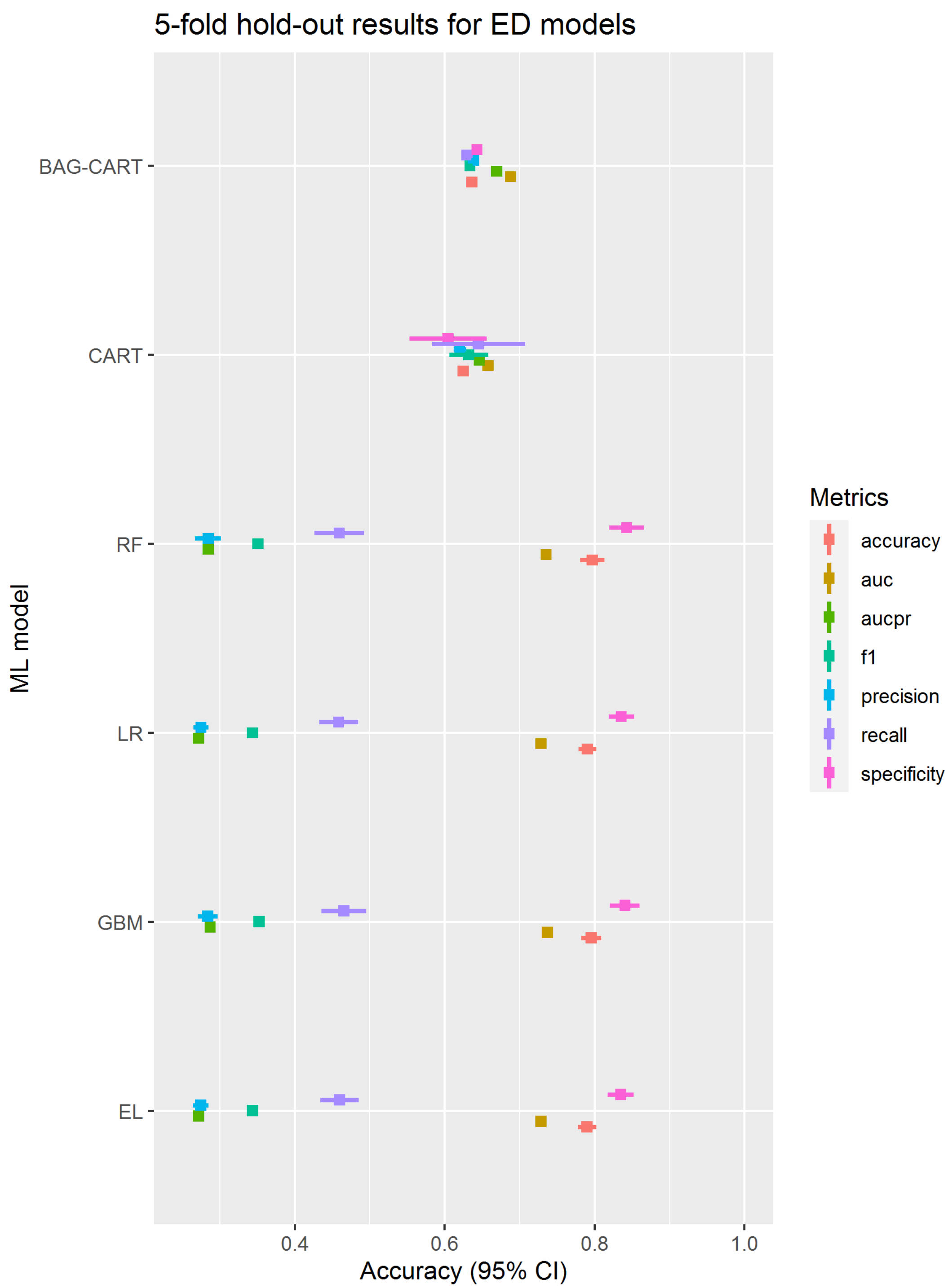


Figure 4.1: Cross-Validation 95% Confidence Interval for ED models

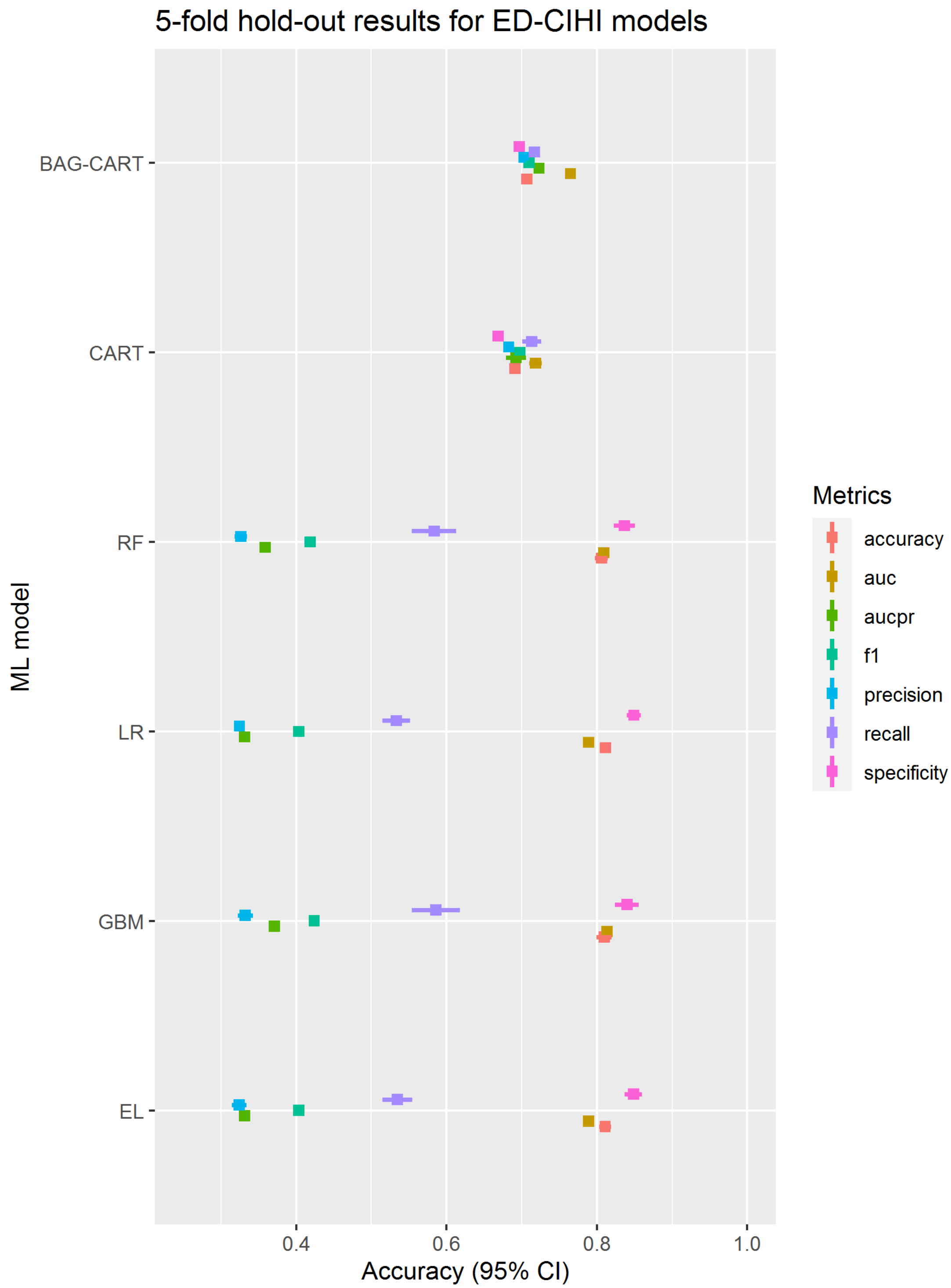


Figure 4.2: Cross-Validation 95% Confidence Interval for in-hospital models

4.4 ROC curves

ROC curves are another good measure of a predictive model's performance. ROC is a graph that shows performance at all classification thresholds. Its parameters are True Positive Rate and False Positive Rate. Figure 4.3 and Figure 4.4 depict ROC curves for all 6 ML algorithms used in this study.

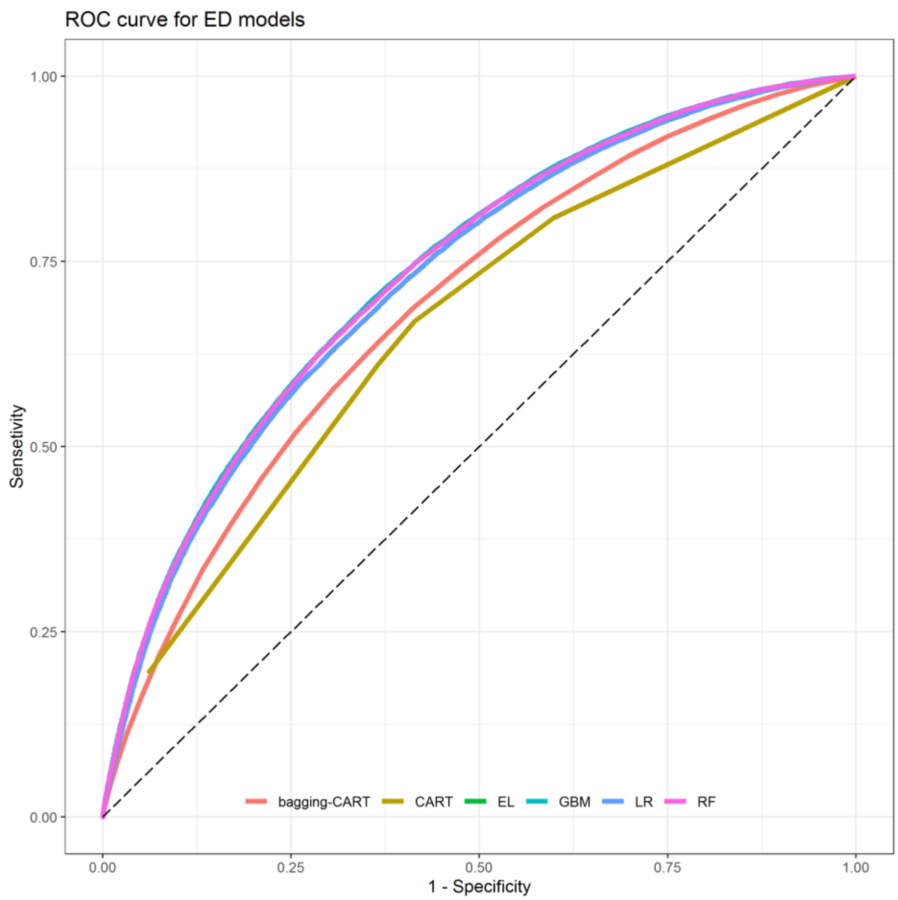


Figure 4.3: ROC curves for ED models

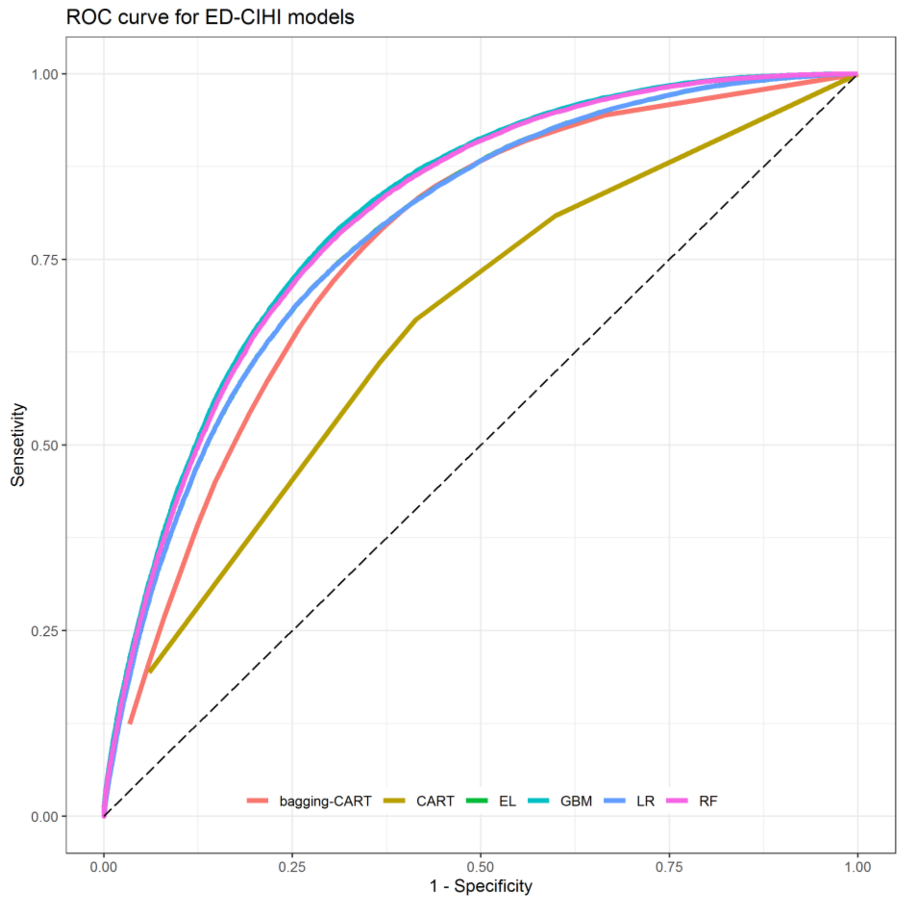


Figure 4.4: ROC curves for in-hospital models

4.5 Variable Importance

Figure 4.5 depicts the scaled importance of the top 20 variables for the top-performing ML algorithm in the first stage (i.e., in-hospital). Appendix B: Variable Importance in ML models provides the full details of variable importance across all ML algorithms. The top predictors of ALC designation for the GBT using predictors from both ED and acute care hospitals were (1) length of stay in acute care; this is the number of days a patient stays in acute beds before being designated as ALC or immediate discharge; (2) method of arrival to ED, walk-in or by an ambulance; (3) main diagnosis being a physical injury (meaning any ICD-10 code starting with the letter “S”); (4) patient’s age and (5) whether the patient stayed in a Special Care Unit (SCU) during their time in the hospital.

Figure 4.6 illustrates the first 20 variables in the GBT in the second stage of the analytics (i.e., using only predictors in ED). The full list and importance of all predictors in the ED stage are in Appendix B: Variable Importance in ML models. Physical injury, method of arrival, and age were again the top predictors of ALC designation. Moreover, the number of investigative technologies that were used on the patient in ED and their types (X-ray, Cat-scan, etc.) were among the top-performing variables. The triage score assigned to patients at ED was one of the top-6 predictors of delayed discharge in this model. We also found that the frailty index (HFRS) was among our top predictors in both time points.

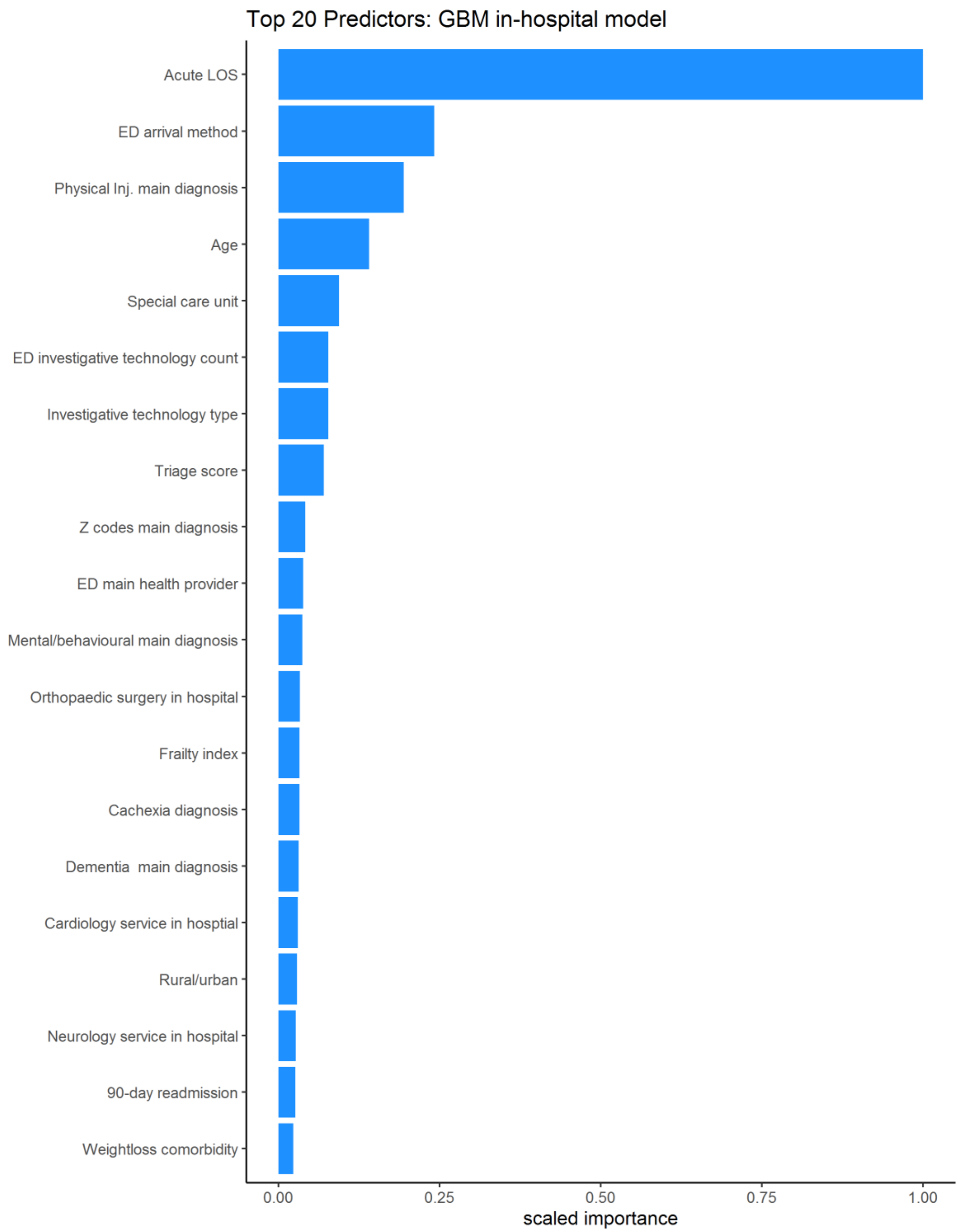


Figure 4.5: Top Predictors of delayed discharge for in-hospital model

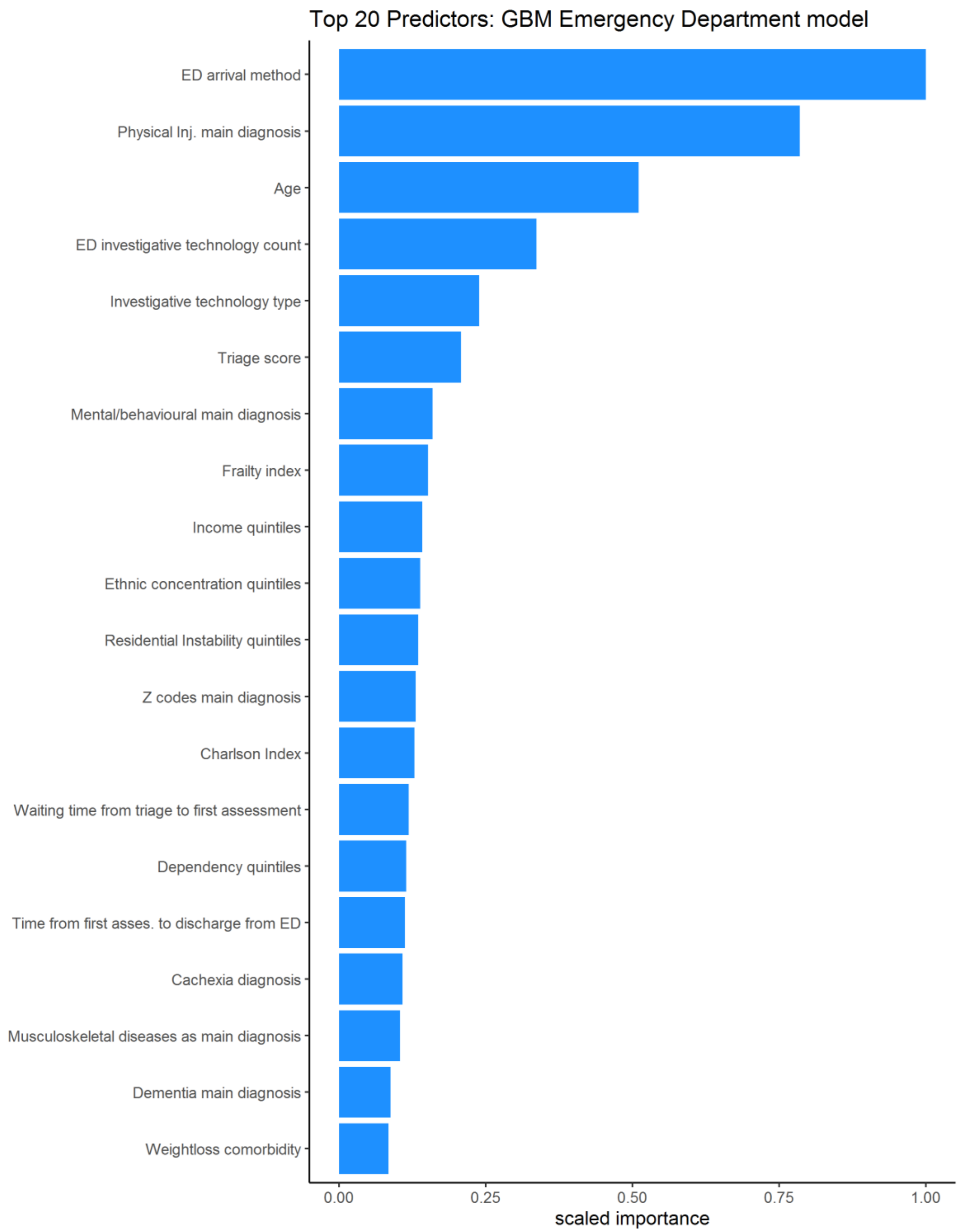


Figure 4.6: Top Predictors of delayed discharge for ED model

5. Discussion

5.1 Summary of Findings

To our knowledge, no study has used machine learning techniques with a sample size as large as us (n = 1,474,285) to predict ALC. We were able to do so long before its occurrence at two different important time points in the patient's trajectory. In the first stage, i.e., in the ED, before the patient is admitted to hospital, we were able to predict ALC with an AUC of 0.74. GBT was the leading ML algorithm across the majority of performance measures. In the second stage, i.e., in the hospital, as we added more information, GBT produced an AUC of 0.81. From ED to hospital, recall was increased by 0.1 and became 0.6 and a specificity of 0.9 was achieved in both time points..

Regarding the performance of ALC prediction, this study reported a higher value of AUC than what was reported in other studies. Arthur *et al.* (2021) achieved an AUC of 0.67 with a smaller sample size (n = 210,931) and Turcotte, Daniel and Hirdes (2020) were able to predict ALC designation with an AUC of 0.76 (n = 30,657). The time point for both studies was in the hospital. The highest prediction accuracy (AUC = 0.85) reported in the literature belongs to Bai *et al.* (2019), however, not only they used a significantly smaller sample size (n = 4,311) and observations were limited to a single hospital, but also they did not use ML. Their outcome measure was not ALC designation as in our study and ROC is obtained using a clinical prediction rule.

In our study, we found that advanced age, female sex, urbanism, diagnosis of dementia, and any other mental/behavioral diagnosis in ED were among the important predictors of ALC. These findings are consistent with prior studies (Challis *et al.*, 2014; Little, 2016; Bai *et al.*, 2019). Interestingly, the method of arrival to ED was the second most important predictor at the in-hospital time point and the most important at ED. More specifically, it suggests that older adults who entered ED via ambulance were most likely to become ALC. Lenzi *et al.* (2014) also found admittance via ambulance to be a major predictor of delayed discharge. Also, our study determined that when a physical injury diagnosis was the main reason for the visit to the first contact point (ED), it becomes an important predictor of becoming ALC. Physical injury flag was a feature based on ICD-10 diagnosis codes in the data.

Another interesting result is that we found that the hospital frailty risk score was a stronger predictor of ALC than the Charlson comorbidity index. The HFRS was in the top 15 features at both stages, whereas the Charlson comorbidity index was not a predictor in the hospital time point and a weaker predictor in ED. Arthur *et al.* (2021) measured the frailty index to be useful in predicting ALC in their initial bivariate analyses. However, it was not a significant predictor in their final model as their data contained other clinical scales that measured subdomains of frailty (e.g., ADL, IADL). As our data did not have such measures, we can conclude that HFRS is a good predictor of ALC in both time points.

Acute length of stay (i.e., the time patient stayed in hospital beds before becoming ALC) was the most informative predictor across all ML algorithms in the hospital. In the literature, similar studies found the length of stay to be a determinant of ALC (Amy *et al.*, 2012; Challis *et al.*, 2014). High length of stay could be an indicator of the complexity and severity of medical needs for older adults staying in acute beds (Amy *et al.*, 2012; Toh *et* *al.*, 2017). After receiving care, such reasons could be still existing, causing their discharge planning to take more time and become ALC. Moreover, staying in special care units while in the hospital was among the top five predictors of ALC in the second time point.

To our knowledge, we are the first study that use clinical variables in ED to predict ALC in hospitals. As a result, we found the number of investigative technologies used for each patient and its type (X-ray, Cat-scan, etc.) were among the top five and top ten predictors of ALC at ED and hospital time points, respectively. These predictors could be indicative of complex medical situations or certain diagnosis groups that require more testing for the main health provider to make clinical decisions (Pines, 2009; Mogensen, Borch and Brandslund, 2011). Moreover, the triage score assigned in ED was also among the top ten predictors at both time points. These findings are valuable in timely prediction of ALC as ED is the first point of contact for most ALC patients (Ahmed, 2019).

Another interesting finding of our study was that patients who received orthopedics surgery service in the hospital were much more likely to become ALC. This finding aligns with the performance of physical injury as a significant predictor. Also, having an orthopedic as a main service provider in ED was among the predictors in our model. These findings could be justified in the way that the aforementioned group suffers from limited mobility and impeded function. Therefore, they may have more post-acute-care needs, such as rehabilitation services, a reasoning that is consistent with the literature (Jerath *et al.*, 2020).

5.2 Clinical and Policy Implications

Integrating data sets from different health sectors and using them to predict ALC designation at different time points has numerous clinical and policy implications. First, identifying ALC patients early on could be helpful to healthcare managers to start their discharge planning and resource allocations sooner. As a result, communications with long-term care, palliative care, or community care facilities could be initiated promptly. Moreover, it could provide policy planners with evidence-based data-driven insights, highlighting the complex needs of ALC patients, leading to better resource planning, as well as optimized training and staffing in LTC or other discharge destinations.

A more integrated approach is key in addressing the problem of ALC. Clinical variables obtained in ED for those who are at risk of becoming ALC can be shared with the care team in the hospital. Helping them to prepare for the complex medical needs of such patients, therefore, reduces the number of ALC days. Gathering information in a timely manner and sharing it among different health sectors (e.g., home-care, ED, acute hospital, and post-acute care facilities) would ensure proactive planning instead of current reactive practices (Arthur *et al.*, 2021). Additionally, families and home support teams would also benefit from knowing the possibility of the patient's ALC occurrence at earlier time points in the patient's trajectory.

As mentioned before, the issue of ALC is costly to the health care system (Home Care Ontario, 2014; Ahmed, 2019). Evidence-based ML processes such as our study could become integrated into clinical decision support tools within the EHR for operationalization purposes. The impact of this practice is better management of ALC cases and saving the health system millions of dollars (Bender and Holyoke, 2018). In general, the cost-saving potential of data-driven decision-making in the case of ALC is significant.

5.3 Strengths, Limitations, and Future Research

5.3.1 Strengths

There were several strengths to this study. First, our sample size was significantly larger than other related works. This would cause better statistical and clinical inference from the data. Helping us draw more accurate conclusions. Moreover, our data were collected from different ED and hospitals across Ontario. This had a positive impact on the generalizability of our results.

Second, our study was the first study to predict ALC at two different time points, providing the decision-makers with more flexibility around the trade-off between time and accuracy. Also, the comparison of the variable importance between the two time points was another strength of our study. Third, in conducting this study, we went beyond common statistical methods (e.g., logistic regression) and used ML on a large data set. Using different ML algorithms enabled us to take advantage of all the data we had.

5.3.2 Limitations

One major limitation of this study was the absence of some useful predictors as administrative datasets have their limitations. In particular, we did not use marital information, caregiver status, living arrangements (living alone, etc.) and source of admission to ED (from home, LTC, etc.). All of these patient-level variables are discussed in the literature as major predictors of delayed discharge (Bai *et al.*, 2019; Turcotte, Daniel and Hirdes, 2020; Arthur *et al.*, 2021). Moreover, we did not have access to common clinical variables and assessments tailored for older adults. Arthur *et al.* (2021) used ADL, IADL, MAPLe, CHESS, and many more variables in their predictive model. These variables were also shown to be contributing to the occurrence of ALC. While we did not have access to these important predictors, we were able to outperform most prior studies.

One problem is the inconsistencies in definitions of ALC across different hospitals due to different hospital policies, physicians and patient flow. Therefore, the results could be biased by some misclassification errors. Looking at differences among hospitals could be a topic of future research.

5.3.3 Future research

Future research should aim to build on the mentioned limitations and focus on improving ML accuracies. This could be done by integrating data sets from other health sectors such as RAI-HC and LTC to include more useful features. Moreover, adding premorbid assessments and pre-existing clinical characteristics of older adults could help implement predictive models at even earlier time points (e.g., upon entry to ED).

Another area of focus could be the addition of organizational factors regarding ALC. Most studies in the literature were based on patient-level characteristics. Adding variables informing of capacities in different discharge destinations and patient flow could also enhance ALC prediction. Future studies could also check for agreement among various ML algorithms on the variable importance and comparing them with the odds ratio from

Logistic Regression. This would show the robustness of the variable importance as well as its comparisons with statistical measures.

One interesting clinical implication using the results of our study is their insights into making decisions about the ALC patient or wait until the next time point (i.e., hospital) to gain more confidence about the predictions. This waiting, however, could be costly. The trade-off between the two could be an interesting topic of future research, which could be carefully analyzed using operation research methods.

Moreover, a limitation of our study is that the predictive accuracies mentioned are average for all patients, some have lower accuracy, and some have higher. It would be interesting to focus on those of higher predictive accuracy in different time points or study the clinical characteristics of both high-performing and low-performing groups for more insights.

6. Conclusion

To our knowledge, this was the first study to predict ALC in two different time points for older adults. We used a series of ML methods to predict ALC designation and compared their results between the two points in patient trajectory. We were able to obtain an AUC score of 0.74 in ED and improve that to 0.81 in the hospital. Furthermore, our study highlighted a number of interesting features that are predictive of ALC. This information could be used to help decision-makers with better resource management and proactive planning for discharge destinations in a timely manner.

Bibliography

Ahmed, D. (2019) The Challenge of Alternate Level of Care (ALC) Facing Older Adults in Ontario: Implications for Government and Policy Makers Using a Descriptive Data Analytics Approach. McMaster University. Available at: http://hdl.handle.net/11375/24951.

Altman, N. and Krzywinski, M. (2018) 'The curse(s) of dimensionality', Nature Methods, 15(6), pp. 399–400. doi:10.1038/s41592-018-0019-x.

Amy, C. et al. (2012) 'Acute care alternate-level-of-care days due to delayed discharge for traumatic and non-traumatic brain injuries', Healthcare Policy = Politiques De Sante, 7(4), pp. 41–55.

Archer, S. (2016) Ontario's Alternate Level of Care (ALC) Problem is Killing Acute Care Hospitals: Time to cut the Gordian knot. | Department of Medicine | School of Medicine | Queen's University. Available at: https://deptmed.queensu.ca/dept-blog/ontarios-alternate-level-care-alc-problem-killing-acute-care-hospitals-time-cut-gordian (Accessed: 30 July 2021).

Arthur, S.A. et al. (2021) 'Do premorbid characteristics of home care clients predict delayed discharges in acute care hospitals: a retrospective cohort study in Ontario and British Columbia, Canada', BMJ Open, 11(2), p. e038484. doi:10.1136/bmjopen-2020-038484.

Bai, A.D. et al. (2019) 'Risk factors, costs and complications of delayed hospital discharge from internal medicine wards at a Canadian academic medical centre: retrospective cohort study', BMC Health Services Research, 19(1), p. 935. doi:10.1186/s12913-019-4760-3.

Bender, D. and Holyoke, P. (2018) 'Why some patients who do not need hospitalization cannot leave: A case study of reviews in 6 Canadian hospitals', Healthcare Management Forum, 31(4), pp. 121–125. doi:10.1177/0840470418755408.

Challis, D. et al. (2014) 'An examination of factors influencing delayed discharge of older people from hospital: Delayed discharge of older people from hospital', International Journal of Geriatric Psychiatry, 29(2), pp. 160–168. doi:10.1002/gps.3983.

Charlson, M.E. et al. (1987) 'A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation', Journal of Chronic Diseases, 40(5), pp. 373–383. doi:10.1016/0021-9681(87)90171-8.

Chawla, N.V. et al. (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', Journal of Artificial Intelligence Research, 16, pp. 321–357. doi:10.1613/jair.953.

CIHI (2009) Alternate level of care in Canada. Canadian Institute for Health Information. Available at: https://secure.cihi.ca/free_products/ALC_AIB_FINAL.pdf.

CIHI (2020) 'National Ambulatory Care Reporting System metadata (NACRS)'. Available at: https://www.cihi.ca/en/national-ambulatory-care-reporting-system-metadata-nacrs.

Claesen, M. and De Moor, B. (2015) ‘Hyperparameter Search in Machine Learning’, arXiv:1502.02127 [cs, stat] [Preprint]. Available at: http://arxiv.org/abs/1502.02127 (Accessed: 31 August 2021).

Costa, A.P. et al. (2012) ‘Acute care inpatients with long-term delayed-discharge: evidence from a Canadian health region’, BMC Health Services Research, 12(1), p. 172. doi:10.1186/1472-6963-12-172.

Domingos, P. (1999) ‘MetaCost: a general method for making classifiers cost-sensitive’, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’99. the fifth ACM SIGKDD international conference, San Diego, California, United States: ACM Press, pp. 155–164. doi:10.1145/312129.312220.

Elixhauser, A. et al. (1998) ‘Comorbidity Measures for Use with Administrative Data’:, Medical Care, 36(1), pp. 8–27. doi:10.1097/00005650-199801000-00004.

Fernandes, M. et al. (2020) ‘Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing’, PLOS ONE. Edited by I. Olier, 15(3), p. e0229331. doi:10.1371/journal.pone.0229331.

Francis, N.K. et al. (2015) ‘The use of artificial neural networks to predict delayed discharge and readmission in enhanced recovery following laparoscopic colorectal cancer surgery’, Techniques in Coloproctology, 19(7), pp. 419–428. doi:10.1007/s10151-015-1319-0.

Friedman, J.H. (2002) ‘Stochastic gradient boosting’, Computational Statistics & Data Analysis, 38(4), pp. 367–378. doi:10.1016/S0167-9473(01)00065-2.

Gilbert, T. et al. (2018) ‘Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study’, The Lancet, 391(10132), pp. 1775–1782. doi:10.1016/S0140-6736(18)30668-8.

Health Quality Ontario (2016) Under Pressure: Emergency department performance in Ontario. Available at: http://www.hqontario.ca/Portals/0/documents/system-performance/under-pressure-report-en.pdf.

Home Care Ontario (2014) ‘Facts & Figures - Publicly Funded Home Care’. Available at: https://www.homecareontario.ca/home-care-services/facts-figures/publiclyfundedhomecare.

Jerath, A. et al. (2020) ‘Delayed discharge after major surgical procedures in Ontario, Canada: a population-based cohort study’, Canadian Medical Association Journal, 192(46), pp. E1440–E1452. doi:10.1503/cmaj.200068.

Krawczyk, B., Woźniak, M. and Schaefer, G. (2014) ‘Cost-sensitive decision tree ensembles for effective imbalanced classification’, Applied Soft Computing, 14, pp. 554–562. doi:10.1016/j.asoc.2013.08.014.

Kuhn, M. (2008) 'Building Predictive Models in R Using the **caret** Package', Journal of Statistical Software, 28(5). doi:10.18637/jss.v028.i05.

Kuhn, M. and Johnson, K. (2013) Applied Predictive Modeling. New York: Springer.

Lavergne, M.R. (2015) Regional variation in alternate level of care (ALC) service use in British Columbia hospitals: An opportunity for intervention? Institute for Health System Transformation & Sustainability. Available at: http://ihsts.ca/wp-content/uploads/2016/06/BC-ALC-Report-Lavergne-2015.pdf.

Lenzi, J. et al. (2014) 'Sociodemographic, clinical and organisational factors associated with delayed hospital discharges: a cross-sectional study', BMC Health Services Research, 14(1), p. 128. doi:10.1186/1472-6963-14-128.

Little, J. (2016) Predictors of Alternate Level of Care within Mental Health Inpatient Units across Ontario. UWSpace. Available at: http://hdl.handle.net/10012/10089.

Milne, V., Petch, J. and Tepper, J. (2017) 'Emergency room overcrowding: causes and cures', healthydebate, 15 June. Available at: https://healthydebate.ca/2017/06/topic/emergency-room-overcrowding/.

Mogensen, C.B., Borch, A. and Brandslund, I. (2011) 'Point of care technology or standard laboratory service in an emergency department: is there a difference in time to action? A randomised trial', Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 19(1), p. 49. doi:10.1186/1757-7241-19-49.

Mowbray, F. et al. (2020) 'Predicting hospital admission for older emergency department patients: Insights from machine learning', International Journal of Medical Informatics, 140, p. 104163. doi:10.1016/j.ijmedinf.2020.104163.

Ogink, P.T. et al. (2019) 'Predicting discharge placement after elective surgery for lumbar spinal stenosis using machine learning methods', European Spine Journal, 28(6), pp. 1433–1440. doi:10.1007/s00586-019-05928-z.

Ontario Hospital Association (2019) A Balanced Approach: The Path to Ending Hallway Medicine for Ontario Patients and Families. Available at: https://www.oha.com/Bulletins/A%20Balanced%20Approach%20-%202019%20Pre-Budget%20Submission.pdf.

Pines, J.M. (2009) 'Trends in the Rates of Radiography Use and Important Diagnoses in Emergency Department Patients With Abdominal Pain', Medical Care, 47(7), pp. 782–786. doi:10.1097/MLR.0b013e31819748e9.

Quan, H. et al. (2005) 'Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data', Medical Care, 43(11), pp. 1130–1139. doi:10.1097/01.mlr.0000182534.19832.83.

R Interface for H2O (2020). Available at: http://www.h2o.ai.

Sills, M.R., Ozkaynak, M. and Jang, H. (2021) 'Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning', International Journal of Medical Informatics, 151, p. 104468. doi:10.1016/j.ijmedinf.2021.104468.

Stock, D. et al. (2016) 'Determinants of alternate-level-of-care delayed discharge among acute care survivors of hypoxic-ischemic brain injury: a population-based cohort study', CMAJ open, 4(4), pp. E689–E697. doi:10.9778/cmajo.20150123.

Sundararajan, V. et al. (2004) 'New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality', Journal of Clinical Epidemiology, 57(12), pp. 1288–1294. doi:10.1016/j.jclinepi.2004.03.012.

Sutherland, J.M. and Crump, R.T. (2013) 'Alternative level of care: Canada's hospital beds, the evidence and options', Healthcare Policy = Politiques De Sante, 9(1), pp. 26–34.

Toh, H.J. et al. (2017) 'Factors associated with prolonged length of stay in older patients', Singapore Medical Journal, 58(3), pp. 134–138. doi:10.11622/smedj.2016158.

Tran, T. et al. (2014) 'A framework for feature extraction from hospital medical data with applications in risk prediction', BMC Bioinformatics, 15(1), p. 425. doi:10.1186/s12859-014-0425-8.

Turcotte, L. and Hirdes, J. (2015) 'Long-stay alternate level of care in Ontario complex continuing care beds', Ontario Hospital Association [Preprint]. Available at: https://www.oha.com/Documents/Long-Stay%20Alternate%20Level%20of%20Care%20in%20Ontario%20Complex%20Continuing%20Care%20Beds.pdf.

Turcotte, L.A., Daniel, I. and Hirdes, J.P. (2020) 'The Post-Acute Delayed Discharge Risk Scale: Derivation and Validation With Ontario Alternate Level of Care Patients in Ontario Complex Continuing Care Hospitals', Journal of the American Medical Directors Association, 21(4), pp. 538-544.e1. doi:https://doi.org/10.1016/j.jamda.2019.12.022.

Victor, C.R. et al. (2000) 'Older patients and delayed discharge from hospital: Older patients and delayed discharge from hospital', Health & Social Care in the Community, 8(6), pp. 443–452. doi:10.1046/j.1365-2524.2000.00270.x.

Walker, J.D., Morris, K. and Frood, J. (2009) 'CIHI survey: Alternative level of care in Canada: a summary', Healthcare Quarterly (Toronto, Ont.), 12(2), pp. 21–23. doi:10.12927/hcq.2009.20674.

Appendices

Appendix A: Descriptive Analysis Table

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
Discharge disposition			<0.001
Acute care	71,280 (5.5%)	3,330 (1.9%)	
Long term care	164,250 (13%)	95,774 (55%)	
Ambulatory/ Palliative care	9,558 (0.7%)	2,869 (1.6%)	
Home with support	359,781 (28%)	47,790 (27%)	
Home without support	570,960 (44%)	7,861 (4.5%)	
Signed out against medical advice	8,052 (0.6%)	276 (0.2%)	
Died	116,080 (8.9%)	16,299 (9.4%)	
Other	96 (<0.1%)	29 (<0.1%)	
Geographic Location			<0.001
Urban	1,056,366 (81%)	151,968 (87%)	
Rural	243,503 (19%)	22,217 (13%)	
(missing)	188	43	
Residential Instability Quintiles			<0.001
1 (least unstable)	153,175 (12%)	16,884 (9.8%)	
2	211,686 (16%)	24,791 (14%)	
3	255,314 (20%)	31,542 (18%)	
4	286,558 (22%)	38,613 (22%)	
5 (most unstable)	377,151 (29%)	60,649 (35%)	
(missing)	16,173	1,749	
Dependency Quintiles			<0.001
1 (least dependent)	148,117 (12%)	19,123 (11%)	
2	186,589 (15%)	24,799 (14%)	
3	225,914 (18%)	29,457 (17%)	
4	274,406 (21%)	35,728 (21%)	
5 (most dependent)	448,858 (35%)	63,372 (37%)	
(missing)	16,173	1,749	
Ethnic Concentration			<0.001

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
Quintiles			
1 (lowest)	324,263 (25%)	37,307 (22%)	
2	284,898 (22%)	37,542 (22%)	
3	234,052 (18%)	33,944 (20%)	
4	216,628 (17%)	32,800 (19%)	
5 (highest)	224,043 (17%)	30,886 (18%)	
(missing)	16,173	1,749	
Income Quintiles			<0.001
1 (lowest)	288,170 (22%)	40,969 (24%)	
2	273,568 (21%)	37,551 (22%)	
3	250,753 (19%)	32,610 (19%)	
4	246,073 (19%)	31,942 (18%)	
5 (highest)	235,058 (18%)	30,504 (18%)	
(missing)	6,435	652	
Gender			<0.001
Male	629,673 (48%)	72,576 (42%)	
Female	670,384 (52%)	101,652 (58%)	
Age Group			<0.001
65-67	296,940 (23%)	22,427 (13%)	
68-73	335,501 (26%)	33,380 (19%)	
74-80	342,127 (26%)	49,642 (28%)	
+81	325,489 (25%)	68,779 (39%)	
Triage			
1 = Resuscitation	65,439 (5.0%)	7,704 (4.4%)	
2 = Emergent	566,091 (44%)	60,476 (35%)	
3 = Urgent	598,923 (46%)	93,836 (54%)	
4 = Less-Urgent (Semi-Urgent)	62,177 (4.8%)	11,050 (6.4%)	
5 = Non-Urgent	4,653 (0.4%)	911 (0.5%)	
(missing)	2,774	251	
Blood Transfused			<0.001
N	1,276,474 (98%)	172,109 (99%)	
Y	23,403 (1.8%)	2,102 (1.2%)	
(missing)	180	17	
Clinical Decision Unit Flag			0.011
N	1,254,059 (96%)	167,854 (96%)	

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
Y	45,998 (3.5%)	6,374 (3.7%)	
Mode of Arrival to ED			<0.001
Walk-in	526,010 (40%)	34,162 (20%)	
Ambulance	774,047 (60%)	140,066 (80%)	
CACS Anesthetic Technique code			<0.001
No Anaesthetic	1,270,444 (98%)	169,677 (97%)	
General/Spinal/Epidural/Neuraxial	191 (<0.1%)	25 (<0.1%)	
Other nerve block/Monitored care	1,594 (0.1%)	223 (0.1%)	
Unmonitored	16,628 (1.3%)	2,566 (1.5%)	
Local	11,020 (0.8%)	1,720 (1.0%)	
(missing)	180	17	
CACS investigative technology category			<0.001
Xray	634,288 (49%)	81,053 (47%)	
Cat Scan	315,946 (24%)	59,032 (34%)	
Not Applicable	302,163 (23%)	29,986 (17%)	
Ultrasound	44,874 (3.5%)	3,632 (2.1%)	
Others	2,786 (0.2%)	525 (0.3%)	
CACS investigative technology category count 1			<0.001
Not Applicable	302,163 (23%)	29,986 (17%)	
1 Intervention within Investigative Technology Category	881,049 (68%)	122,408 (70%)	
2 or more Interventions within Investigative Technology Category	116,464 (9.0%)	21,794 (13%)	
(missing)	381	40	
CACS investigative technology total count			
0	302,163 (23%)	29,986 (17%)	
1	718,355 (55%)	78,516 (45%)	
2	236,439 (18%)	51,448 (30%)	
3	36,055 (2.8%)	11,516 (6.6%)	
4	5,832 (0.4%)	2,265 (1.3%)	
5	891 (<0.1%)	401 (0.2%)	
6	134 (<0.1%)	71 (<0.1%)	
7	8 (<0.1%)	7 (<0.1%)	

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
8	0 (0%)	1 (<0.1%)	
(missing)	180	17	
CACS Partition			<0.001
Emergency Visit (EV) Acute Admission Partition	1,272,912 (98%)	171,478 (98%)	
Diagnosis Partition	26,144 (2.0%)	2,667 (1.5%)	
Intervention Partition	821 (<0.1%)	66 (<0.1%)	
(missing)	180	17	
Glasgow coma scale			<0.001
Not a coma patient	1,292,735 (99%)	172,848 (99%)	
Any coma scale	7,322 (0.6%)	1,380 (0.8%)	
Main Service Provider in ED			<0.001
General Practitioner	369,604 (28%)	40,888 (23%)	
Emergency Medicine	764,482 (59%)	111,231 (64%)	
Others	165,786 (13%)	22,089 (13%)	
(missing)	185	20	
Orthopedics main provider	11,138 (0.9%)	4,206 (2.4%)	<0.001
30-day revisit to ED	107,690 (8.3%)	14,662 (8.4%)	0.061
Hospital Frailty Risk Score			<0.001
Less than 5	77,692 (6.0%)	21,674 (12%)	
>= 5	1,222,365 (94%)	152,554 (88%)	
Cachexia in main diagnoses or as comorbidity			<0.001
No	1,288,169 (99%)	167,900 (96%)	
Yes	11,888 (0.9%)	6,328 (3.6%)	
Triage to first assessment by physician time quartile			**<0.001**
1st	316,598 (26%)	36,291 (22%)	
2nd	311,153 (25%)	40,950 (25%)	
3rd	302,408 (25%)	42,747 (26%)	
4th	299,321 (24%)	46,898 (28%)	
(missing)	70,577	7,342	
Assessment to left ED time quartile			**<0.001**
1st	313,964 (25%)	37,095 (22%)	
2nd	307,423 (25%)	41,787 (25%)	
3rd	306,737 (25%)	42,244 (25%)	

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
4th	303,789 (25%)	45,850 (27%)	
(missing)	68,144	7,252	
Acute Length of Stay median	6.45 (8.19)	13.01 (16.80)	<0.001
Prior 90-day hospitalization			**<0.001**
0	959,740 (74%)	126,949 (73%)	
1	340,317 (26%)	47,279 (27%)	
Prior 365-day hospitalization			**0.002**
0	711,894 (55%)	94,722 (54%)	
1	588,163 (45%)	79,506 (46%)	
90-day readmission			**<0.001**
0	954,709 (73%)	135,019 (77%)	
1	345,348 (27%)	39,209 (23%)	
Specialized Clinical Interventions Received in Hospital			
Cardioversion			**<0.001**
0	1,294,025 (100%)	173,835 (100%)	
1	6,032 (0.5%)	393 (0.2%)	
Cell Saver			**<0.001**
0	1,298,074 (100%)	174,058 (100%)	
1	1,983 (0.2%)	170 (<0.1%)	
Chemotherapy			**<0.001**
0	1,294,522 (100%)	173,207 (99%)	
1	5,535 (0.4%)	1,021 (0.6%)	
Dialysis			**<0.001**
0	1,274,988 (98%)	170,308 (98%)	
1	25,069 (1.9%)	3,920 (2.2%)	
Tube Feeding			**<0.001**
0	1,293,002 (99%)	170,967 (98%)	
1	7,055 (0.5%)	3,261 (1.9%)	
Heart Resuscitation			**<0.001**
0	1,291,780 (99%)	173,575 (100%)	
1	8,277 (0.6%)	653 (0.4%)	
Mechanical Ventilation (long term)			**<0.001**
0	1,285,281 (99%)	170,642 (98%)	

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
1	14,776 (1.1%)	3,586 (2.1%)	
Mechanical Ventilation (short term)			**<0.001**
0	1,261,401 (97%)	169,974 (98%)	
1	38,656 (3.0%)	4,254 (2.4%)	
Parenteral Nutrition			**<0.001**
0	1,291,027 (99%)	172,252 (99%)	
1	9,030 (0.7%)	1,976 (1.1%)	
Paracentesis			**<0.001**
0	1,289,398 (99%)	172,446 (99%)	
1	10,659 (0.8%)	1,782 (1.0%)	
Pleurocentesis			**<0.001**
0	1,278,101 (98%)	170,649 (98%)	
1	21,956 (1.7%)	3,579 (2.1%)	
Radiotherapy			<0.001
0	1,294,612 (100%)	172,592 (99%)	
1	5,445 (0.4%)	1,636 (0.9%)	
Tracheostomy			<0.001
0	1,296,911 (100%)	172,945 (99%)	
1	3,146 (0.2%)	1,283 (0.7%)	
Vascular Access Device			<0.001
0	1,239,115 (95%)	162,006 (93%)	
1	60,942 (4.7%)	12,222 (7.0%)	
Flagged Intervention Count			<0.001
0	1,086,952 (84%)	141,960 (81%)	
1	82,496 (6.3%)	11,985 (6.9%)	
2	54,781 (4.2%)	8,444 (4.8%)	
3	55,939 (4.3%)	7,605 (4.4%)	
4	4,751 (0.4%)	1,253 (0.7%)	
5	1,240 (<0.1%)	344 (0.2%)	
6	11,386 (0.9%)	1,995 (1.1%)	
7	2,394 (0.2%)	631 (0.4%)	
8	118 (<0.1%)	11 (<0.1%)	

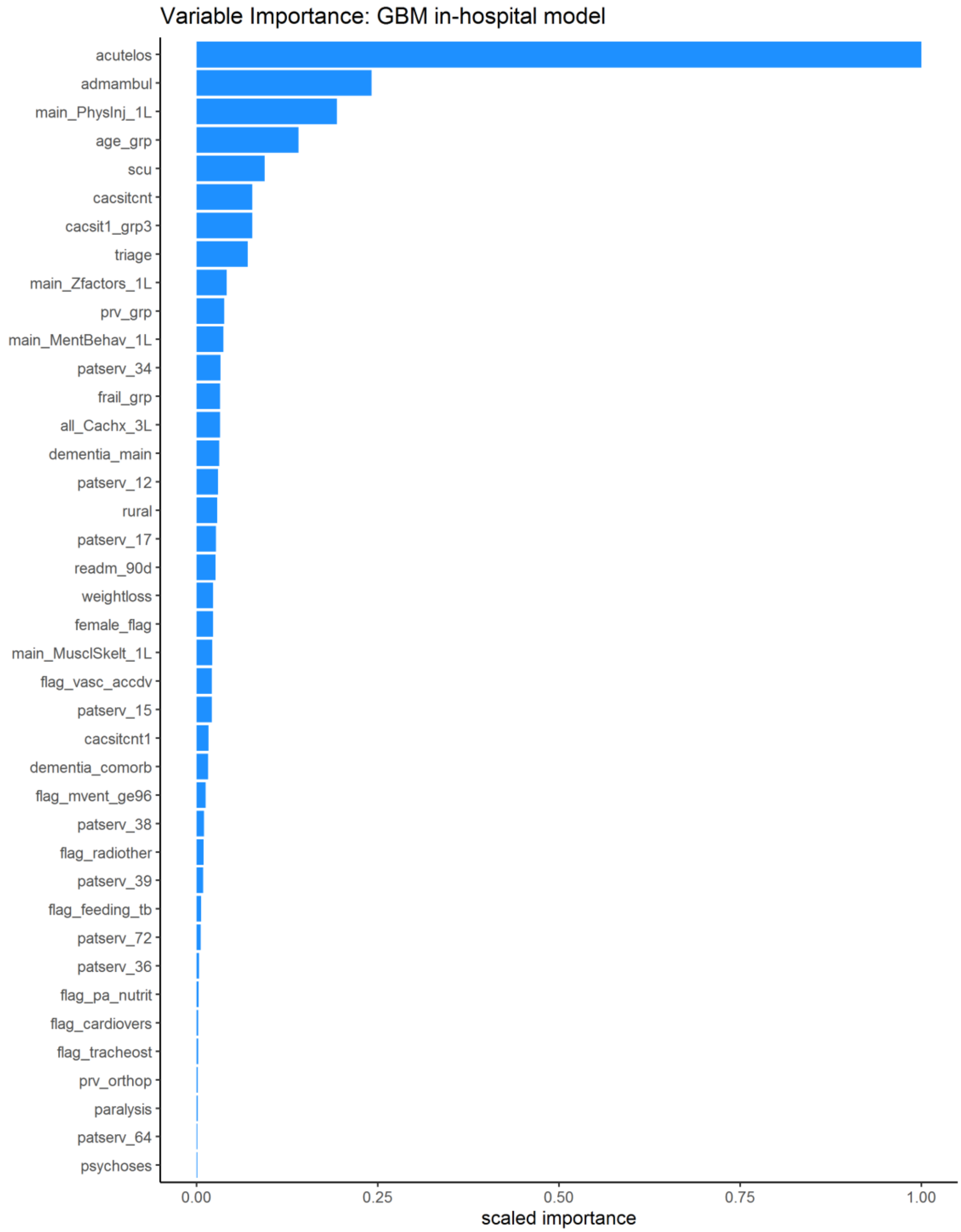
Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
Special Care Unit			
Medical Intensive Care Nursing Unit	17,726 (1.4%)	1,460 (0.8%)	
Surgical Intensive Care Nursing Unit	4,666 (0.4%)	1,257 (0.7%)	
Trauma Intensive Care Nursing Unit	63 (<0.1%)	5 (<0.1%)	
Combined Medical/Surgical Intensive Care Nursing Unit	119,310 (9.2%)	14,586 (8.4%)	
Burn Intensive Care Nursing Unit	416 (<0.1%)	83 (<0.1%)	
Cardiac Intensive Care Nursing Unit Surgery	4,066 (0.3%)	448 (0.3%)	
Coronary Intensive Care Nursing Unit Medical	46,293 (3.6%)	3,077 (1.8%)	
Neonatal Intensive Care Nursing Unit	33(<0.1%)	3(<0.1%)	
Neurosurgery Intensive Care Nursing Unit	4,496 (0.3%)	1,312 (0.8%)	
Respirology Intensive Care Nursing Unit	97 (<0.1%)	23 (<0.1%)	
Step-Down Medical Unit	10,668 (0.8%)	1,779 (1.0%)	
Combined Medical/Surgical Step-Down Unit	16,328 (1.3%)	2,006 (1.2%)	
Step-Down Surgical Unit	5,673 (0.4%)	871 (0.5%)	
No SCU	1,070,222 (82%)	147,318 (85%)	
Charlson comorbidities:			
Dementia (comorbidity)			<0.001
0	1,283,915 (99%)	168,647 (97%)	
1	16,142 (1.2%)	5,581 (3.2%)	
Dementia (main diagnosis)			<0.001
0	1,296,028 (100%)	170,539 (98%)	
1	4,029 (0.3%)	3,689 (2.1%)	
Diabetes with complications			<0.001
0	1,138,950 (88%)	155,428 (89%)	
1	161,107 (12%)	18,800 (11%)	
Cerebrovascular disease			<0.001
0	1,290,652 (99%)	172,515 (99%)	
1	9,405 (0.7%)	1,713 (1.0%)	
Elixhauser comorbidities:			

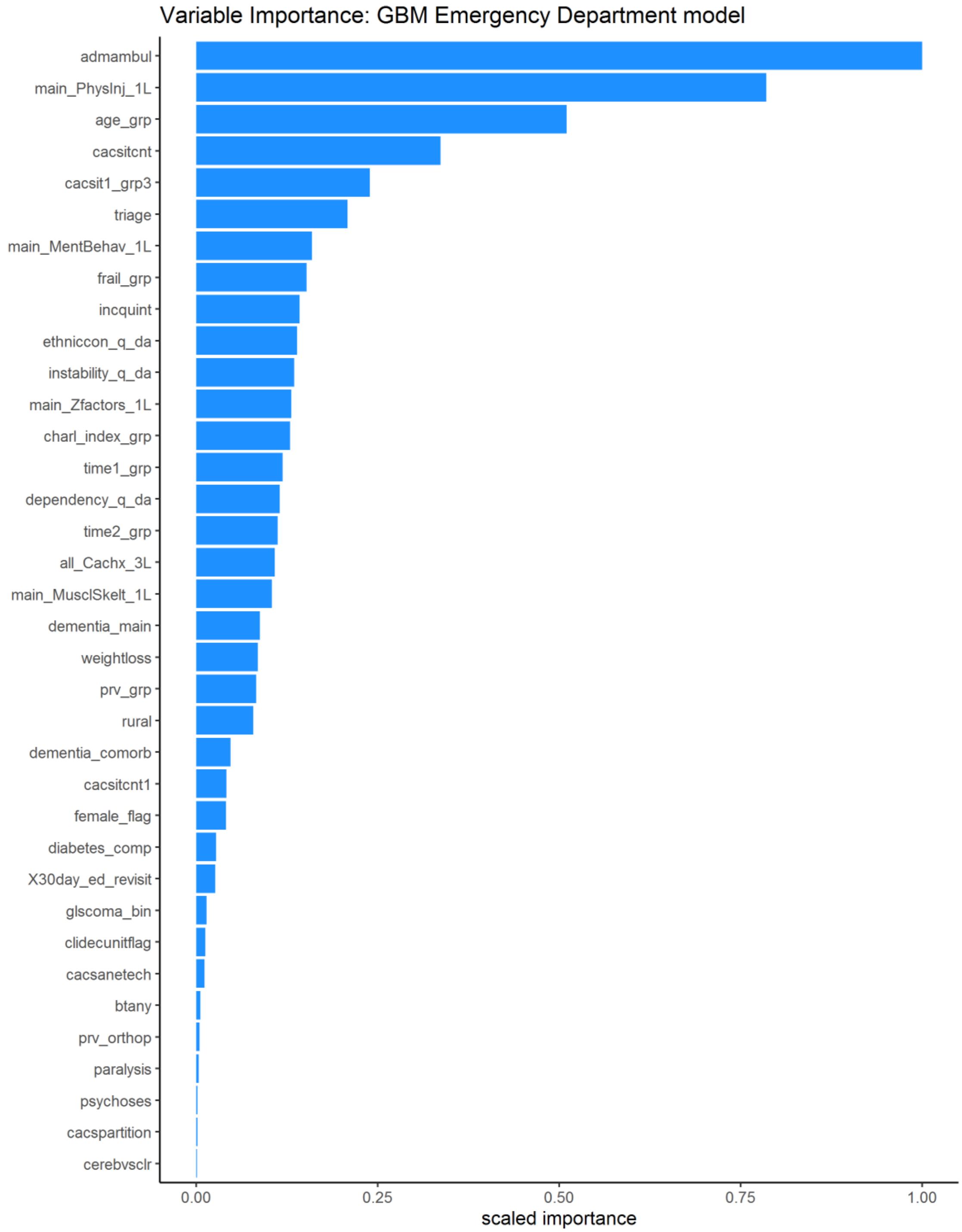
Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
Paralysis			<0.001
0	1,298,330 (100%)	173,761 (100%)	
1	1,727 (0.1%)	467 (0.3%)	
Weight loss			<0.001
0	1,286,201 (99%)	167,556 (96%)	
1	13,856 (1.1%)	6,672 (3.8%)	
Psychoses			<0.001
0	1,298,819 (100%)	173,848 (100%)	
1	1,238 (<0.1%)	380 (0.2%)	
Charlson Index			<0.001
0	689,597 (53%)	94,568 (54%)	
1	313,234 (24%)	42,538 (24%)	
2+	297,226 (23%)	37,122 (21%)	
Main diagnosis group:			
Physical Injuries (S)			<0.001
0	1,204,466 (93%)	136,408 (78%)	
1	95,591 (7.4%)	37,820 (22%)	
Mental and Behavioral (F)			<0.001
0	1,280,960 (99%)	166,454 (96%)	
1	19,097 (1.5%)	7,774 (4.5%)	
Factors influencing health status and contact with health services (Z)			<0.001
0	1,286,256 (99%)	169,210 (97%)	
1	13,801 (1.1%)	5,018 (2.9%)	
Musculoskeletal (M)			<0.001
0	1,272,965 (98%)	167,155 (96%)	
1	27,092 (2.1%)	7,073 (4.1%)	
Main patient service in acute care:			
Orthopaedic Surgery			<0.001
0	1,255,492 (97%)	154,212 (89%)	
1	44,565 (3.4%)	20,016 (11%)	
Geriatrics			<0.001
0	1,297,358 (100%)	173,209 (99%)	
1	2,699 (0.2%)	1,019 (0.6%)	
Traumatology			<0.001

Characteristic	**Non-ALC Patients**[1] (N = 1,300,057)	**ALC patients**[1] (N = 174,228)	**p-value**[2]
0	1,280,800 (99%)	167,639 (96%)	
1	19,257 (1.5%)	6,589 (3.8%)	
Neurology			<0.001
0	1,277,163 (98%)	167,655 (96%)	
1	22,894 (1.8%)	6,573 (3.8%)	
Rheumatology			<0.001
0	1,298,874 (100%)	173,893 (100%)	
1	1,183 (<0.1%)	335 (0.2%)	
Psychiatry			<0.001
0	1,298,543 (100%)	173,814 (100%)	
1	1,514 (0.1%)	414 (0.2%)	
Plastic Surgery			<0.001
0	1,299,086 (100%)	174,169 (100%)	
1	971 (<0.1%)	59 (<0.1%)	
Urology			<0.001
0	1,270,527 (98%)	172,470 (99%)	
1	29,530 (2.3%)	1,758 (1.0%)	
Gynaecology			<0.001
0	1,298,016 (100%)	174,115 (100%)	
1	2,041 (0.2%)	113 (<0.1%)	
Cardiology			<0.001
0	1,172,012 (90%)	167,302 (96%)	
1	128,045 (9.8%)	6,926 (4.0%)	
Gastro-Enterology			<0.001
0	1,248,533 (96%)	171,517 (98%)	
1	51,524 (4.0%)	2,711 (1.6%)	
Thoracic Surgery			<0.001
0	1,297,503 (100%)	174,104 (100%)	
1	2,554 (0.2%)	124 (<0.1%)	

[1]Frequency (%) or Median (IQR) [2]Wilcoxon rank sum test; Pearson's Chi-squared test

Appendix B: Variable Importance in ML models




Variable Importance: GBM Emergency Department model
admambul
main_PhysInj_1L
age_grp
cacsitcnt
cacsit1_grp3
triage
main_MentBehav_1L
frail_grp
incquint
ethniccon_q_da
instability_q_da
main_Zfactors_1L
charl_index_grp
time1_grp
dependency_q_da
time2_grp
all_Cachx_3L
main_MusclSkelt_1L
dementia_main
weightloss
prv_grp
rural
dementia_comorb
cacsitcnt1
female_flag
diabetes_comp
X30day_ed_revisit
glscoma_bin
clidecunitflag
cacsanetech
btany
prv_orthop
paralysis
psychoses
cacspartition
cerebvsclr
0.00
0.25
0.50
0.75
1.00
scaled importance