Analysing repeated hospital readmissions using data mining techniques

Ofir Ben-Assuli^a and Rema Padman^b

^aFaculty of Business Administration, Information Systems Department, Ono Academic College, Kiryat Ono, Israel; ^bThe H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

Few studies have examined how to identify future readmission of patients with a large number of repeat emergency department (ED) visits. We explore 30-day readmission risk prediction using Microsoft's AZURE machine learning software and compare five classification methods: Logistic Regression, Boosted Decision Trees (BDTs), Support Vector Machine (SVM), Bayes Point Machine (BPM), and Two-Class Neural Network (TCNN). We predict the last readmission visit of frequent ED patients extracted from the electronic health records of their 8455 penultimate visits. The methods show differential improvement, with the BDT indicating marginally better AUC (area under the ROC curve) than logistic regression and BPM, followed by the TCNN and SVM. A comparison of BDT and Logistic Regression results for correct and incorrect classification highlights the similarities and differences in the significant predictors identified by each method. Future research may incorporate time-varying covariates to identify other longitudinal factors that can lead to readmission risk reduction.

ARTICLE HISTORY

Received 6 March 2017
Revised 12 September 2017
Accepted 5 October 2017

KEYWORDS

Hospital readmissions; predicting readmissions; logistic regression; boosted decision tree; support vector machine; two-class neural network

1. Introduction

Predicting who is likely to be readmitted and understanding the key factors contributing to preventable readmissions in hospitals is being widely researched, but few studies have examined how to identify future readmissions of patients with a large number of repeat hospital visits (Bardhan, Oh, Zheng, & Kirksey, 2014). Insights regarding the prediction of readmission risks may help physicians and hospital management to develop early interventions, to identify and treat patients at high risk of readmissions (Hagland, 2011), and to enhance both care quality and financial outcomes (Kociol et al., 2012).

Studies have compared data mining techniques applied to complex problems in the health care arena (Dreiseitl et al., 2001; Koh & Tan, 2011), with a limited number exploring data mining techniques for readmission risk prediction (Caruana et al., 2015; Futoma, Morris, & Lucas, 2015; Leeds et al., 2017). The leading IS Health care scholars recommend predictive analytics as one of the two most timely research avenues for IS researchers (Kohli & Tan, 2016).

Logistic regression or logit regression is the most commonly used method implemented in research designed to investigate readmissions including prediction models (Bardhan et al., 2014). The vast majority of the readmission literature has used logistic regression models to estimate the readmission probability of patients (Muus et al., 2010; Philbin, Dec, Jenkins, & DiSalvo, 2001;

Shelton, Sager, & Schraeder, 2000; Silverstein, Qin, Mercer, Fong, & Haydar, 2008). Previous studies mainly account for the first readmission and most previous studies *do not* follow multiple visits for the same person over time (Alexander, Grumbach, Remy, Rowell, & Massie, 1999; Amarasingham et al., 2010; Krumholz et al., 2000; Silverstein et al., 2008).

Artetxe, Beristain, Graña, and Besga (2016) compared two supervised classification models: namely, Support Vector Machines (SVMs) and decision forests to evaluate the risk of readmission within 30 days after discharge. After training a model on various scenarios including heart failure, chronic obstructive pulmonary disease and diabetes mellitus, they found that SVM exhibited the greatest sensitivity and outperformed the decision forest algorithm (Artetxe et al., 2016). Boosted decision trees (BDTs) have only recently started being used in predicting early readmission (Caruana et al., 2015; Futoma et al., 2015).

However, to the best of our knowledge, few comparisons of these data mining techniques have been reported for readmission prediction in the health care arena. In the context of heart failure patients, and specifically congestive heart failure, Hon, Pereira, Sushmita, Teredesai, and De Cock (2016), compared the performance of decision trees, boosted decision trees, and logistic regression models. They suggested that these models could identify patients with an anticipated lengthy and expensive 30-day readmission. They found that logistic regression

was competitive with (and perhaps slightly superior to) the BDT. The least accurate classification model was the regular decision tree.

In this study, we compared the prediction assignment for 30-day readmission (dichotomous variable) of five classification methods: logistic regressions, SVM, the two-class BDT, BPM, and Two-class Neural Network. Given that most studies have not followed multiple visits for the same person over time, we apply these methods to study the problem of readmission risk for patients who presented in the emergency department (ED) at least seven times in the previous four-year period.

We accessed the electronic health record data of 8455 patients with 71,816 visits to the ED (each patient had between seven and twelve visits to the ED). We analysed up to the penultimate visit of each of the 8,455 patients to predict the last visit's readmission outcome (readmission/no readmission) using Microsoft AZURE Machine Learning (AZURE ML) software.

This paper is organised as follows: Section 2 presents an overview of the literature. Section 3 introduces the data and the classification methods and their related literature as they apply to the health care IT arena. Section 4 reports the research findings. Section 5 discusses the results and the contributions of this study. It also suggests future research avenues to extend this work and resolve some of its limitations.

2. Literature survey

2.1. Predictive analytics in the health care arena

Advances in health information technologies have prompted many health care providers to consider prediction analytics (Thayer, Bruno, & Remorenko, 2013). The health care sector is increasingly looking for smarter, more informed, decision-making capability to improve health outcomes and provide more efficient health care services to patients. Rapid evolution of clinical analytics that seeks to analyse the vast amounts of clinical information generated during the encounter with and management of the patient by the provider, and obtaining new insights from these studies, is a critical component of modern big data analytics (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014).

2.2. Predictive analytics for readmissions

One of the main promising targets for prediction analytics is readmission reduction. Getting the rate of readmissions down has become the goal of virtually all health care organisations, clinicians, insurers, hospital managers, and policy-makers (Schneider et al., 2012; Stone & Hoffman, 2010) to the point that recent policies may even penalise providers and their organisations for high readmission rates (Cress, 2011). Wang and Hajli (2017) described the value of analytical approaches to

parallel-process large data volumes that capture patients' full data from medical records. These approaches can identify previously unnoticed patterns in patient data related to hospital readmissions (Wang & Hajli, 2017). However, predictive analytics requires high-quality data on valid predictors in order to generate credible results and useful insights. Attribute selection methods allow feature engineering that facilitate this process.

2.3. Attribute selection methods

Data mining applications are often associated with data engineering as a central issue of the researcher. Trustworthy classifications of predictive feature sets result in reliable data models that in their turn ensure credible learning schemes. The incorporation of inconsequent and wrong characteristics in the model building process may result in poor prediction execution. Feature choice requires a combination of search and attribute value assessment in addition to consideration of specific learning schemes. Hall and Holmes (2003) introduce a prototype of a few attribute selection systems which present features ranking. Feature selection is performed by cross-validating the rankings and considering a learning layout to discover the optimal characteristics.

Liu and Yu (2005) present other feature selection algorithms, review current characteristics of algorithms for categorisation and grouping, collect and compare various algorithms within a framework of investigation schemes, assessment norms and data mining, offer standards in selecting characteristics, and propose a consolidated platform. They suggest that unrelated or unnecessary characteristics (eg, age) should be eliminated before classification algorithm performance is evaluated, so that they do not delay operational implementation. In order to identify and eliminate these attributes as well as to improve the efficiency of clustering, they use correlation analysis or feature selection method. However, one of the disadvantages of this method is that a researcher might miss a valuable relationship between a set of independent variables and a dependent variable. An example supporting this concept is presented by Yoo et al. (2012), who mention that an individual who smokes, drinks, and has *Helicobacter pylori* infection may likely be excluded from further investigation of stomach cancer, however, all of these attributes together may be substantially connected to the cancer. Thus, these noise reduction methods should be applied carefully.

2.4. Classification models

A few studies have begun to compare data mining techniques in readmission prediction (Caruana et al., 2015; Futoma et al., 2015; Leeds et al., 2017). Caruana et al. (2015) discussed the trade-off between accuracy and intelligibility in data mining models. Models such

as boosted decision trees, random forests, and neural nets are accurate but typically not intelligible or easily interpreted and understood. However, intelligible models such as logistic regression, Naive-Bayes, and single decision trees often have significantly worse accuracy. The authors also described the application of a learning method based on high-performance generalised additive models to predicting 30-day readmissions for pneumonia.

Futoma et al. (2015) compared models for predicting early hospital readmissions (30 day readmissions). They trained deep neural networks and compared them to penalised logistic regressions. Their results show that the neural networks significantly outperformed the penalised logistic regressions. Leeds et al. (2017) compared various logit regression models to identify significant predictors of readmission. Time-varying and fixed data for all patients were analysed using a time-to-event regression model to identify significant time-point predictors of discharge on a given hospital day. Twenty-two daily clinical measures were significant in both regression models. Both models demonstrated good discrimination ability on the comparison of discharge behaviours after complex surgery. IS researchers are often asked to contribute to the design of algorithms by modelling and evaluating prediction analytics techniques to enrich EHR use in analytics (Kohli & Tan, 2016). These studies require access to data across the whole health care network to achieve better clinical outcomes (Fraser et al., 2013).

Models of patient-level factors such as medical comorbidities, basic demographic data, and clinical variables are much better predictors of mortality than readmission risk (Amarasingham et al., 2010; Hammill et al., 2011; van Walraven et al., 2010). The usefulness of wider social, environmental, and medical factors that contribute to readmission risk has not been widely studied (Kansagara et al., 2011). Bardhan et al. (2014), for instance, found that HIT usage, patient demographics, visit features, insurance type, and hospital features were all related to patient readmission risk in the context of congestive heart failure patients.

3. Data and methods

3.1 Data

This study is based on data from a large HMO in the State of Israel. This HMO deployed the EHR system analysed here to serve over 3.8 million patients. Utilising the nation-wide Health Information Exchange (HIE) network, the system retrieves historical patient data from many medical applications at the HMO's hospitals and clinics. This data retrieval architecture provides an integrated, real-time, virtual patient record that is available at all points of care delivery of the HMO.

In this regard, Israel's Ministry of Health has deployed a programme for HMOs that implements creative

empirical and theoretical work to reduce readmissions, including the extensive use of electronic health information technologies. Physicians already access patients' medical records through a HIE. Policies stipulating concrete steps to prevent readmissions are also being formulated. However, despite these efforts, readmission rates remain unacceptably high (Haklai et al., 2013; Meydan, Haklai, Gordon, Mendlovic, & Afek, 2015). In the US in 2011, 20 percent of Medicare patients were readmitted to hospitals within 30 days (Hagland, 2011).

We extracted data utilising the HIE network from the EHR systems of several EDs across Israel. In this study, we targeted patients who are potentially at higher risk of readmissions based on the pattern of their frequent hospital visits and associated readmissions. The final data-set consisted of 8455 patient records (chronic and non-chronic). Specifically, we studied patients who had between 7 and 12 visits over a duration of 4 years, which provided sufficient data from a sizeable population to extract insights. There is limited research on frequently readmitted patients to help define this concept. A 2014 study (Black, 2014) defined frequent patients specifically for readmission purposes as having 3–5 visits within a time period of about a year and half to two years. Another study defined frequent readmissions for patients having three ED visits a year or more (Kirby, Dennis, Jayasinghe, & Harris, 2010). With our data extract across four years, we thus found it reasonable to define frequent patients as those having between 7 to 12 visits to the EDs, and this definition was validated with senior physicians. Additionally, we determined that changing this number to a minimum of five visits within four years produces similar results.

For each of the 8455 patients, we collected relevant, available information up to the penultimate visit, and predict the readmission result associated with the last visit (whether the last visit of each patient ended in readmission or not). The data included patients' demographics, permanent medications, adverse reactions, detailed lab and imaging results, past diagnoses, health care procedures, and so on.

3.2 Dependent and independent variables

3.2.1. Dependent variable – readmission

Quantified whether the patient was readmitted on the last visit within 30 days after being discharged from the hospital on the penultimate visit (coded 1) or not (coded 0).

This measure is widely used as a means of monitoring the efficacy of critical care pathways (Ramachandran, Erraguntla, Mayer, & Benjamin, 2007). An accepted notion is that the shorter the period between discharge and readmission, the more likely that the patient was discharged prematurely (Ather, Chung, Gregory, & Demissie, 2004). Nevertheless, readmission rates are also used as a proxy for quality of care rendered during the

hospitalisation (Ather et al., 2004; Welch, Larson, Hart, & Rosenblatt, 1992). In addition, it reflects the level of patient safety as part of the health care service as a whole.

3.2.2. Main independent variables

3.2.2.1. EHR use. The term “EHR use” refers to access to at least one of several medical history components in the EHR utilising the HIE network. This was measured as a dichotomous variable in the following way: 1 – for each patient for whom the EHR was used at least once across all visits up to and including the penultimate one, and 0 – otherwise.

Vest (2009) found that system access is not random, and that specific patient factors increased the likelihood of information access. His findings indicated that the more a patient's data were examined, the more likely that this individual had more ED visits and inpatient hospitalisations.

3.2.2.2. Health maintenance organisation/insurance.

The EHR chosen for this study provides integrated information only on patients belonging to the main HMO. On other patients, only prior information from the same hospital as the current referral is available. To control for major discrepancies in the quality and the amount of medical information between the main HMO patients and other HMO patients, a dichotomous variable was created as follows: 1 – if the patient was a member of the main HMO on the penultimate visit, for whom full medical history was available via the EHR, or 0 – if the patient was not a member of the main HMO on the penultimate visit.

3.2.2.3. ED Unit. This variable represented the specific type of unit where the patient was evaluated in the ED on the penultimate visit. The ED units in the log-file were internal medicine, surgery, obstetrics, orthopaedics, gynaecology, ENT (ear, nose & throat), primary, and dermatology.

3.2.2.4. Hospital. This variable represented the specific hospital where the patient was evaluated on the penultimate visit. We included this variable to account for potential variations across hospitals, such as policies. Each one of the hospitals was assigned a different code represented by this variable.

3.2.2.5. Creatinine result. A continuous variable represented the result of the Serum Creatinine test during the penultimate visit; namely, the level of Serum Creatinine in the blood. This was the result of the Creatinine test ordered by the physician to determine whether the result was “normal” or not. This variable was measured in milligrammes per decilitre. In addition to the penultimate result, we also included the standard deviation of the results of the Creatinine test across all previous visits from the first visit to the penultimate visit.

3.2.2.6. Differential diagnosis. This variable represents the diagnosis made by the physician for each case (according to diagnostic standards of the ICD9) (Ben-Assuli, Shabtai, & Leshno, 2013; Ben-Assuli, Sagi, Leshno, Ironi, & Ziv, 2015; Kachalia et al., 2007). We created dichotomous variables for the seven most frequent diagnoses including heart disease, nephritis UTI, obstructive pulmonary disease, blood forming organs, pneumonia, enteritis and colitis and cerebrovascular disease. For each one of these variables, we assigned the value of 1 for each patient for whom the differential diagnosis was assigned at least once in all visits up to the penultimate one, and 0 otherwise.

3.2.2.7. LOS – length of stay. Measured in days and expressing the actual admission days (only for admitted patients) on the penultimate visit (Horn et al., 1991; Kelly, Sharp, Dwane, Kelleher, & Comber, 2012; Marang-van de Mheen, Lingsma, Middleton, Kievit, & Steyerberg, 2014; Mayer, Yaron, & Lowenstein, 2010).

3.2.2.8. Days between Referrals. Measured the interval in days between the penultimate and the antepenultimate visits. This variable has been examined in previous studies as well (Frankl, Breeling, & Goldman, 1991; Lindrooth & Weisbrod, 2007; Spratt, Finkelstein, Butler, Badger, & Crowley, 1987; Yang, Delcher, Shenkman, & Ranka, 2016). In addition, we also included the average interval in days between every two consecutive visits, from the first visit until the penultimate visit.

Other independent variables were age, gender, total number of ED visits and total number of previous readmissions (up to the penultimate visit). None of the variables included any information from the final visit. Correlation analysis of all the variables using Pearson correlation indicated that there were no highly correlated variables in this data-set.

3.3. Methodology of binary classification models

We use Microsoft's AZURE ML platform (Barga, Fontama, Tok, & Cabrera-Cordon, 2015) to conduct our experiments. Figure 1 depicts the main steps in the data mining comparison process using AZURE ML including:

- (1) Extracting data from EHR, collecting all the data into the data-sets, and separating the data into columns.
- (2) Clean missing data (data preparation) – Decisions were needed regarding wrong or missing data. We decided whether to fix the wrong data or complete the missing data (by changing it/adding the mean value).
- (3) Select columns in data-set – Select the relevant columns in the data-sets by selecting the columns used in the model from the entire database.

Figure 1. Two-class classification comparisons with AZURE ML: BDT, logistic regression, SVM, BPM and two-class neural network (TCNN).

- (4) Split data – splitting the data into two parts for training and testing the models, respectively. About 70% of the data was used to train the models. The other 30% was used to evaluate the models. This partitioning and data-sets were identical for all five classification models tested.
- (5) Attributes selection – The module of feature selection in Azure applies well-known statistical methods to the variables provided as input, creating metrics that can be used to identify the columns that have the best information value. In our study, we differentiate the variables that have impact on our target variables from the ones that have no impact on our target variables. Consequently, the non-impacting variables were selected out via this step.
- (6) Training model – Separate training models were executed for each one of the five classification models. Training a classification model in our supervised machine learning demands that the data-set should contain both the outcome that we tried to predict (regarding readmissions), and all related factors (variables). The train

model uses 70% of the data-set to extract statistical patterns and build a model (the other 30% of the data-set are remaining to test the models).

This model includes one main preliminary step: Parameter optimisation – This module performs a parameter sweep on each models to determine the optimum parameter settings. Generally, it is used to build and test models using different combinations of settings, in order to determine the optimum parameters for the given prediction task and data (for instance, in the BDT, the number of trees constructed and maximum number of leaves per tree).

- (7) Score model – This module tested the prediction accuracy for all the trained classification models on the 30% of the data-set that have not been used for training the models.
- (8) Evaluate model – This AZURE module is used to compare the results of all five models (using R programming to merge all results), create the receiver operating characteristic (ROC) curves and the comparative measurement tables.

3.4. Attribute selection in AZURE ML

Azure ML presents multiple attribute selection methods to deal with all data types. Additionally, as a part of a learning process several ML algorithms apply some kind of feature selection or dimensionality reduction.

The feature selection modules in Azure ML operates with familiar statistical methods to the data columns provided as input, producing metrics that can be used to recognise the columns that have the best information importance. We differentiate the variables that have impact on our target variables from the ones that have no impact on our target variables. Consequently, the non-impacting variables were removed via this step.

We performed the attribute selection using AZURE ML in our five tested models as can be viewed in Figure 1.

3.5. Classifiers descriptions

3.5.1. Method 1: logistic regression

We first explored logistic regression modelling and analysis, the most popular method in prior research and has been implemented in many readmission studies (Bardhan et al., 2014; Ben-Assuli et al., 2013; Ben-Assuli, Shabtai, Leshno, & Hill, 2014).

Our model used a binary dependent variable, y_i , for each patient i , for each variable k , representing readmission occurrence (1 = readmission; 0 = non-readmission). The model is specified as:

$$P(y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon)} \quad (1)$$

Defining $\pi_i = P(y_i = 1)$ and $1 - \pi_i = P(y_i = 0)$, we have

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 \cdot \text{Insurance} + \beta_2 \cdot \text{Age} + \beta_3 \cdot \text{Gender} + \beta_4 \cdot \text{Creatinine Result} + \dots + \epsilon \quad (2)$$

The ratio $\frac{\pi_i}{1 - \pi_i} = \frac{P(y_i=1)}{P(y_i=0)}$ represents the odds of the event $y_i = 1$ occurring (ϵ represents the random error). Therefore, as commonly reported, the odds ratio (OR) is presented in the results in Table 2.

3.5.2. Method 2: two-class boosted decision tree

A two-class BDT (Freund & Schapire, 1995; Witten, Frank, Hall, & Pal, 2016) is an ensemble learning method in which predictions are based on an entire ensemble of trees that correct for errors together. It is suitable for a binary dependent variable such as the readmission decision. In this method, forecasts are established on the entire set of trees that performs the forecast.

A previous study showed superior results with the BDT compared to other decision trees including decision forest, bagging decision tree, and randomised decision tree (Dietterich, 2000). BDT can improve accuracy at the cost of refraining from predicting training cases that are very tough to classify (Drucker & Cortes, 1995; Freund & Mason, 1999).

BDTs have only recently started to be used in predicting early readmission (Caruana et al., 2015; Futoma et al., 2015). Stiglic et al. (2015) conducted comprehensive predictive modelling for the evolution of medical complications using paediatric hospitals' discharge data. They used regularised logistic regression and comorbidity-based features in the first step, and BDTs in the final step of their analysis. Their results showed improvement in the comprehensibility of the final predictive model but the authors also mentioned that it was difficult to interpret the models since each BDT had to be interpreted separately rather than simply merging all rules obtained during the process of decision tree boosting.

Figure 2 presents an example of one BDT produced by AZURE ML (out of 479 trees).

Figure 2. Example of one BDT produced by AZURE ML (out of 479 trees).

3.5.3. Method 3: two-class SVM

Support vector machines (SVMs) are supervised learning models used for classification tasks (Cortes & Vapnik, 1995; Teo, 2012) to recognise patterns from large volumes of data. This classifier is useful for predicting binary outcomes that depend on continuous and categorical independent variables.

Given a set of labelled training examples with binary outcome values, the SVM algorithm assigns new examples into one category or the other such that the two categories are divided by the widest gap possible. New examples are then predicted to belong to a category based on which side of the gap they are identified with. There are many successful applications of SVM models ranging from information retrieval to text and image classification. SVMs have rarely been used to predict readmissions. Rumshisky et al. (2016) used SVM to predict early psychiatric readmission and found that using SVM would facilitate interventions and reduce the risk and related readmission costs.

Our SVM model was trained using the retrospective data on hospital visits up to the 11th visit and tested on its prediction of the 12th visit. It is calibrated using parameter optimisation of AZURE ML. For instance, in the SVM, we can calibrate various parameters: Number of iterations (denotes the number of iterations used when building the model) and Lambda (weight for L1 regularisation).

3.5.4. Method 4: two-class bayes point machine

The Naïve Point Machine is implemented in AZURE ML via the Bayes Point Machine. The Bayes Point Machine (Herbrich, Graepel, & Campbell, 2001) is a Bayesian approach to linear classification. They (Herbrich et al., 2001) present two algorithms to stochastically approximate the centre of mass of version space: a billiard sampling algorithm and a sampling algorithm based on the well known perceptron algorithm. It is shown how both algorithms can be extended to allow for soft-boundaries in order to admit training errors. Training errors happen when your training data are noisy or if your model is too simple. They also demonstrate that the real-valued output of single Bayes points on novel test points is a valid confidence measure and leads to a steady decrease in generalisation error when used as a rejection criterion.

It efficiently approximates the theoretically optimal Bayesian average of linear classifiers (in terms of generalisation performance) by choosing one “average” classifier, the Bayes Point. Because the Bayes Point Machine is a Bayesian classification model, it is not prone to overfitting to the training data. It does not require parameter sweeping nor data to be normalised.

BPMs have rarely been used to predict readmissions. We found two studies of the same crew predicting readmission within a year. The first paper (Mesgarpour, Chaussalet, & Chahed, 2016) achieved AUC ranging from 73% to 74.3%. The second study used BPM method

to predict the emergency readmission to NHS hospitals (Mesgarpour, Chaussalet, & Chahed, 2017). They chose BPM, since it is not prone to overfitting, highly efficient in approximating the Bayesian average classifier (Mesgarpour et al., 2017) improving their AUC from their previous work (Mesgarpour et al., 2016) reported above to 77.1%.

3.5.5. Method 5: two-class neural network

A neural network is a set of interconnected layers, in which the inputs lead to outputs by a series of weighted edges and nodes. The weights on the edges are learned when the neural network is trained on the input data. The direction of the graph proceeds from the inputs through the hidden layer, with all nodes of the graph connected by the weighted edges to nodes in the next layer.

Most predictive tasks can be accomplished easily with only one or a few hidden layers. Recent research has shown that deep neural networks (DNN) can be very effective in complex tasks such as image or speech recognition, in which successive layers are used to model increasing levels of semantic depth.

To compute the output of the network for any given input, a value is calculated for each node in the hidden layers and in the output layer. For each node, the value is set by calculating the weighted sum of the values of the nodes in the previous layer and applying an activation function to that weighted sum.

As in previous studies, we used a TCNN (which is a supervised learning method) to create a model that predicts a target variable that has only two values (He, Wang, Graco, & Hawkins, 1997; Ou & Murphey, 2007). The standard Neural network has been used before to predict readmission (eg, Zheng et al., 2015), but to the best of our knowledge, there is no study that has incorporated the Two-Class Neural Network to predict readmission, though Tsai et al. (2016) utilised the regular neural network as a two-class approach.

3.6. AZURE Machine Learning (ML) platform

We used the Microsoft AZURE machine learning software (Barga et al., 2015) to analyse the data-set using logistic regression, SVM, BDT, TCNN, and Bayes Point Machine classification models. AZURE ML Studio is a GUI-based integrated development environment for constructing and operationalising machine-learning workflows. The basic computational unit of an Azure ML Studio is a module which implements machine-learning algorithms, data conversion and transformation functions, etc. The modules can be connected by data flows, thus implementing a machine-learning pipeline. We added some open source R programming in order to incorporate specific features in each algorithm such as merging files (as shown in Figure 1).

4. Results

We present four sets of results in this study. The first summarises some useful descriptive statistics associated with the 8455 patients that were analysed. The second shows the attribute selection results, and, the third illustrates results from a basic logistic regression model similar to many prior studies. The fourth shows the comparison across all five-classification models and summarises the results for the all classifiers.

4.1. Descriptive statistics

Table 1 provides a brief summary of the key features concerning the patients and their visits to the EDs. Males were significantly older, had higher creatinine levels, higher EHR use and higher 30-day readmissions. Females had a significantly longer average number of days between consecutive hospital visits. Differences in length of stay and insurance coverage were not significant.

4.2. Attribute selection results

We performed Feature Selection using AZURE ML as described in section 3.4, resulting in the selection of a subset of variables. The variables that were found to be

not relevant included: primary care type of ED unit and cerebrovascular disease diagnoses. Consequently, the Classifiers utilised the remaining set of variables.

4.3. Logistic regression classifier

We ran the logistic regression model using all 8455 patient data. The results for predicting 30-day readmissions are shown in Table 2. They indicate that when the age of the patient increases by one year, the likelihood of readmission *increases* by 1.1% (95% CI = 1.006–1.016, OR = 1.011). For males, the likelihood of readmissions within 30 days to the ED *increases* by 16.5% in comparison to females (95% CI = 1.017–1.335, OR = 1.165). For patients whose EHR was accessed during their previous visits, the likelihood of readmissions within 30 days to the ED *increases* by 50.7% in comparison to patients whose EHR was not accessed (95% CI = 1.171–1.94, OR = 1.507). In addition, when mean days between consecutive visits (until the final visit) increase by one day, the likelihood of readmission *decreases* by 0.3% (95% CI = 0.996–0.999, OR = 0.997). When the LOS increases by one day, the likelihood of readmission *increases* by 1.2% (95% CI = 1.001–1.023, OR = 0.997). When the number of prior visits increases by one visit, the likelihood of readmission *decreases* by 22.3% (95%

Table 1. Characteristics of the sample and comparison between male and female subjects across 8455 patients.

Data characteristics	Study sample $n = 8455$	Male $n = 4118$ (48.7%)	Female $n = 4337$ (51.3%)	Sig.
Age	64.8 ± 19.5	65.4 ± 18.5	64.3 ± 20.4	0.007
Insurance (main HMO %)	89.4%	89%	90%	0.27
Creatinine result	1.38 ± 1.35	1.57 ± 1.45	1.2 ± 1.21	<0.001
EHR Use	43.9%	44.3%	43%	0.033
Length of Stay	5.13 ± 6.64	5.15 ± 0.5	5.11 ± 0.5	0.57
Days between referrals	115.5 ± 141.2	112.78 ± 141	118.1 ± 140.7	<0.001
Average number of 30-day readmissions	2.5 ± 1.64	2.58 ± 1.65	2.43 ± 1.63	<0.001

Note: Data are the mean ($\pm$ SD) or number of subjects (proportion).

Table 2. Logistic regression results using all 8455 patient data.

Variables in the model	B	SE	Wald	Sig.	OR	95.0% CI for OR	
						Lower	Upper
Age***	0.011	0.002	19.179	0.000	1.011	1.006	1.016
Gender*	0.153	0.070	4.826	0.028	1.165	1.017	1.335
Insurance	−0.099	0.108	0.830	0.362	0.906	0.732	1.121
Creatinine Result	0.010	0.034	0.080	0.777	1.010	0.944	1.080
PriorCreatinineResults SD	0.150	0.095	2.489	0.115	1.162	0.964	1.401
EHR Use**	0.410	0.129	10.121	0.001	1.507	1.171	1.940
MeanDaysBetweenReferrals**	−0.003	0.001	8.740	0.003	0.997	0.996	0.999
LOS*	0.012	0.005	4.562	0.033	1.012	1.001	1.023
Prior#Visits***	−0.253	0.026	96.768	0.000	0.777	0.739	0.817
PriorReAdmission30Days_Sum***	0.655	0.030	475.54	0.000	1.925	1.815	2.041
HeartDisease	−0.012	0.075	0.026	0.872	0.988	0.853	1.144
NephritisUTI	0.005	0.086	0.003	0.957	1.005	0.848	1.190
COPD	−0.061	0.112	0.299	0.585	0.941	0.756	1.171
BloodFormingOrgans	0.107	0.105	1.031	0.310	1.113	0.905	1.367
Pneumonia	0.079	0.101	0.606	0.436	1.082	0.887	1.320
EnteritisAndColitis	−0.098	0.133	0.546	0.460	0.906	0.699	1.176
CerebrovascularDisease	0.009	0.116	0.006	0.939	1.009	0.804	1.265
Constant	−21.9	27.210	0.000	0.999	0.000		1.016

* $p < 0.05$, + $p < 0.1$.

** $p < 0.01$.

*** $p < 0.001$.

CI = 0.739–0.817, OR = 0.777). Finally, when the number of prior readmission occurrences increases by one, the likelihood of next readmission *increases* by 92.5% (95% CI = 1.815–2.041, adjusted OR = 1.925).

The remaining variables were not significant in predicting readmission within 30 days. The results were adjusted to account for many confounders, including the comorbidities of the patients, 11 ED units and 7 hospitals, which we do not present (but were entered as dichotomous variables).

The remaining variables were not significant in predicting readmission within 30 days. The results were adjusted to account for many confounders, including the comorbidities of the patients, 11 ED units and 7 hospitals, which we do not present (but were entered as dichotomous variables).

4.4. Comparing the five classifiers

This section reports the results of the comparison across all five-classification models. Since the data were divided randomly into the training set (70% of the data) and the testing/validation set (30% of the data), the results shown here report the results on the test data after using parameter optimisation. Appendix A summarises and explains the terminology used in this section (Appendix B shows the list of abbreviations).

Before we made the comparison of all five models, we tested the difference for each one of our five classifiers the results of using all the past readmission history of a patient (full model) vs. just the last visit state information (last visit model). There is a significant difference in the AUC (DeLong, DeLong, & Clarke-Pearson, 1988) in favour of the full model, using all the past readmission history of the patients (Table 3).

Figure 3 shows the ROC curves for the five models. The ROC of the BDT is the highest for any false positive (FP) and true positive (TP) rates, followed by that of the BPM and the logistic regression. The ROC of the TCNN and the SVM methods are the lowest of the five models, and is much lower than the other two models. Similarly, the highest AUC was found for the BDT and the lowest for the SVM (AUC: Decision tree = 0.92, BPM = 0.913, Logistic regression = 0.912, TCNN = 0.878 and SVM = 0.846). After testing the difference between the AUC of each pair of classifier models (DeLong et al., 1988), we found that the BDT outperforms BPM and logistic regression but with borderline significance ($p < 0.1$), however all these three classifiers outperform significantly the TCNN and all four models significantly outperform the SVM (DeLong et al., 1988).

The results in Table 4 show that all five methods performed well when predicting the next readmission within 30 days. The BDT classification model had the highest accuracy (85.3%), with the highest area under the ROC curve (AUC = 92%) and the highest value for true positive prediction of readmissions (788 TP) but

with the lowest true negative predictions (1375 TN). The second and third best classification models were BPM (AUC = 91.3%; Accuracy = 82.4%) and logistic regression (AUC = 91.2%; Accuracy = 81.6%), respectively, having the highest value of true negative prediction of readmissions (logistic regression = 1426 TN and BPM = 1414 TN) and with equivalent precision as the decision tree (~76%). The fourth and the fifth best classification models were the TCNN (AUC = 87.8%; Accuracy = 80.1%) and SVM (AUC = 84.6%; Accuracy = 76.9%), respectively, that had the lowest AUC and accuracy compared to the first three models (which also demonstrated good performance).

Table 4 compares performance across all five methods. As noted, the results showed that all five methods exhibited good performance for predicting the next readmission to hospitals within 30 days. Interestingly, the TP capabilities were better for the BDT but the TN were almost the worst in this model (the worst was TCNN).

The power of the positive predictive value was the same for the decision tree, BPM and logistic regression (precision = ~76%) three of which were superior to SVM (precision = 70.5%) and TCNN (precision = 66.7%). However, the decision tree and the TCNN models were superior to the other models in positively predicting readmissions out of the entire future readmission with a recall of 86.7 and 88.8%, respectively. These two measurements led to the dominance of the BDT for the F1Score (the harmonic mean of precision and recall).

Additional insights regarding comparison of the classifiers for patients that were predicted very good (true positives and negatives) vs. those that were predicted very poorly (false positives and negatives) is available in Appendix C.

Figure 3. Comparisons of ROC curves.

Table 3. Comparing all past visits history of patients (full model) vs. the last visit information (last visit model).

Model	AUC full model	AUC last visit model	Accuracy full model	Accuracy last visit model	Precision full model	Precision last visit model	Recall full model	Recall last visit model	F1 score full model	F1 score last visit model
Boosted Decision Tree	0.92	0.59	0.853	0.642	0.758	0	0.867	0	0.809	0
Logistic Regression	0.912	0.6	0.816	0.646	0.762	0.578	0.707	0.04	0.734	0.076
SVM	0.846	0.57	0.769	0.642	0.705	1	0.613	0.0001	0.656	0.002
BPM	0.913	0.6	0.824	0.646	0.760	0.553	0.743	0.057	0.751	0.104
Two-Class NN	0.878	0.59	0.801	0.6	0.667	0.443	0.888	0.442	0.762	0.443

5. Discussion and conclusion

Much of the readmission literature has primarily used logistic regression models to estimate patients' readmission probability (Bardhan et al., 2014; Muus et al., 2010; Philbin et al., 2001; Shelton et al., 2000; Silverstein et al., 2008). As we report in this paper, logistic regression is a stable method and classifier (accuracy exceeding 80%) in predicting readmission within 30 days. In addition, logistic regression is intelligible (Caruana et al., (2015) and it is easy to interpret the strength of each variable

added several critical diagnoses for each patient and used five distinct classification models for analysing readmission risk for patients with frequent visits to the ED. In addition, as compared to previous works, we include aggregate variables from all visits up to the penultimate visit for each patient, and we predict the final visit's readmission status (whether the last visit of each patient ended as readmission or not). The results from all the classifiers demonstrate clear benefits from using the full medical history and computing aggregate

Table 4. Readmissions prediction results summary for the five models.

Model	AUC	True positive	True negative	False positive	False negative	Accuracy	Precision	Recall	F1 Score
Boosted decision tree	0.925	788	1375	252	121	0.853	0.758	0.867	0.809
Logistic regression	0.912	643	1426	201	266	0.816	0.762	0.707	0.734
SVM	0.846	557	1394	233	352	0.769	0.705	0.613	0.656
BPM	0.913	675	1414	213	234	0.824	0.760	0.743	0.751
Two-class NN:	0.878	807	1224	403	102	0.801	0.667	0.888	0.762

(via the odds ratio, confidence intervals, and significance level).

This study extends previous studies by following the same patient over multiple visits instead of predicting based on a single visit. Most of the previous studies have used data only from the most recent visit in order to predict readmission (Felker et al., 2004; Khanna, Boyle, & Good, 2014). Studies have also attempted to simply explain the statistical connections between different variables and the readmission outcome (Boulding, Glickman, Manary, Schulman, & Staelin, 2011; Donzé, Lipsitz, Bates, & Schnipper, 2013; Kossovsky et al., 1999; Krumholz et al., 2000). Some have included clinical variables associated with some common primary diagnoses related to hospital readmissions.

Studies that have used data from multiple visits of the same patients rather than only the most recent visit (Almagro et al., 2006; Daly, Douglas, Kelley, O'Toole, & Montenegro, 2005; Liu et al., 2015) have mostly not extended their analysis to measure predictive performance. The only study that we found that integrated information from many visits and also did predictive analytics did it specifically for Congestive Heart Failure (CHF) (Bardhan et al., 2014). In the current study, we

variables rather than only using the most recent visit information.

Logistic regression model, BPM and the BDT models display good predictive capabilities for 30-day readmission rates (the decision tree was slightly superior to the other models), outperforming the prediction capability of SVM and Neural Network. Overall, the BDT model was superior to the other models in terms of accuracy and AUC measurements.

We suggest that researchers in health care informatics arena who implement prediction models (especially classification models) should consider evaluating and incorporating more than one classifier. Specifically, we suggest incorporating BDT classifiers and comparing its contribution to other common classifiers (eg, logistic regression). Future research will follow multiple visits for the same patients over time in the prediction model while taking care of potential statistical issues when performing longitudinal analyses (the regression errors are not random when dealing with the same patient over time). These same methods could be used for predictive purposes in other clinical settings as well.

From an application perspective, with the increasing deployment and use of EHRs in the hospitals and the ED, in Israel, US and many other parts of the world, data are

now available to investigate readmissions reduction, a major challenge everywhere. The variables and data we use in this study are representative of this setting, and drawn from several EDs in Israel, not just one. Hence, the approaches we have tested and insights from the analyses should be replicable and generalisable in other similar settings.

On the methodology side, the discussion on False Negative (FN) comparisons between the models is especially useful because not recognising a correct readmission and falsely classifying a non-readmission have differing but significant penalties associated with them. Neural Networks and decision tree have the lowest FNs which is the best. Minimising the FN is the major valuable target in actual practice. The FP comparison, on the other hand, means incorrectly classifying as readmission and admitting or treating a patient, which also has a cost or penalty, but perhaps not as critical as FN. The models and their significant variables that illustrate these contexts are also generalisable to other decisions and settings.

While readmission reduction is a widely studied problem, little is known regarding which models work better, particularly for frequently hospitalised patients that we study in this paper, as well as models that go beyond logistic regression. Hence, a study that compares many predictive modelling approaches, including the BDT that provides the best result, is both informative and generalisable.

The study has some limitations. One limitation has to do with the choice of five (albeit well-known) out of the many data mining classification models. The second possible limitation is that in our EHR data, we only considered one biomarker value (the creatinine test result). In addition, we acknowledge that our findings related to our specific frequent ED visitors and might not be replicable for other sets of patients or points of care. Future research could incorporate other data mining classification methods along with time-varying covariates to identify a more insightful set of longitudinal factors related to readmissions while attempting to predict future readmission events that integrate more biomarkers and current laboratory results. Another possible limitation is that this paper does not describe a causal relationship between the independent variables and our dependent (target) variable. This was not the original purpose of the study, but certainly it is advisable to do so in future research efforts.

Funding

This research was supported in part by the United States-Israel Binational Science Foundation (BSF) [grant number 2015153].

References

- Alexander, M., Grumbach, K., Remy, L., Rowell, R., & Massie, B. M. (1999). Congestive heart failure hospitalizations and survival in California: Patterns according to race/ethnicity. *American Heart Journal*, 137(5), 919–927.
- Almagro, P., Barreiro, B., Ochoa de Echagüen, A., Quintana, S., Rodríguez Carballeira, M., Heredia, J. L., & Garau, J. (2006). Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration*, 73(3), 311–317.
- Amarasingham, R., Moore, B. J., Tabak, Y. P., Drazner, M. H., Clark, C. A., Zhang, S., ... Halm, E. A. (2010). An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Medical Care*, 48(11), 981–988.
- Artetxe, A., Beristain, A., Graña, M., & Besga, A. (2016). Predicting 30-day emergency readmission risk. In *European Transnational Education* (pp. 3–12). Springer.
- Ather, S., Chung, K. D., Gregory, P., & Demissie, K. (2004). The association between hospital readmission and insurance provider among adults with asthma. *Journal of Asthma*, 41(7), 709–713.
- Bardhan, I. R., Oh, J. H., Zheng, Z., & Kirksey, K. (2014). Predictive analytics for readmission of patients with congestive heart failure. *Information Systems Research*, 26(1), 19–39.
- Barga, R., Fontana, V., Tok, W. H., & Cabrera-Cordon, L. (2015). *Predictive Analytics with Microsoft Azure Machine Learning*. Springer.
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33(7), 1123–1131.
- Ben-Assuli, O., Shabtai, I., & Leshno, M. (2013). The impact of EHR and HIE on reducing avoidable admissions: Controlling main differential diagnoses. *BMC Medical Informatics and Decision Making*, 13(1), 541.
- Ben-Assuli, O., Shabtai, I., Leshno, M., & Hill, S. (2014). EHR in emergency rooms: Exploring the effect of key information components on main complaints. *Journal of Medical Systems*, 38(4), 1–8.
- Ben-Assuli, O., Sagi, D., Leshno, M., Ironi, A., & Ziv, A. (2015). Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study. *Journal of Biomedical Informatics*, 55, 31–40.
- Black, J. T. (2014). Learning about 30-day readmissions from patients with repeated hospitalizations. *The American Journal of Managed Care*, 20(6), e200–e207.
- Boulding, W., Glickman, S. W., Manary, M. P., Schulman, K. A., & Staelin, R. (2011). Relationship between patient satisfaction with inpatient care and hospital readmission within 30 Days. *The American Journal of Managed Care*, 17(1), 41–48.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1721–1730). ACM.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.
- Cress, J. C. (2011). Helping reduce hospital readmissions using seven key elements. *Geriatr Care Manage J*, 21, 25–28.

- Daly, B. J., Douglas, S. L., Kelley, C. G., O'Toole, E., & Montenegro, H. (2005). Trial of a disease management program to reduce hospital readmissions of the chronically critically ill. *Chest*, 128(2), 507–517.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837–845.
- Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), 139–157.
- Donzé, J., Lipsitz, S., Bates, D. W., & Schnipper, J. L. (2013). Causes and patterns of readmissions in patients with common comorbidities: Retrospective cohort study. *BMJ*, 347, f7171.
- Dreiseitl, S., Ohno-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H., & Binder, M. (2001). A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics*, 34(1), 28–36.
- Drucker, H., & Cortes, C. (1995). Boosting decision trees. In *Proceedings of the 8th International Conference on Neural Information Processing Systems* (pp. 479–485). MIT Press.
- Felker, G. M., Leimberger, J. D., Califf, R. M., Cuffe, M. S., Massie, B. M., Adams, K. F., ... O'Connor, C. M. (2004). Risk stratification after hospitalization for decompensated heart failure. *Journal of Cardiac Failure*, 10(6), 460–466.
- Frankl, S. E., Breeling, J. L., & Goldman, L. (1991). Preventability of emergent hospital readmission. *The American Journal of Medicine*, 90(6), 667–674.
- Fraser, I. H., Jayadewa, C., Goodwyn, J., Mooiweer, P., Gordon, D., & Piccone, J. (2013). *Analytics across the ecosystem. A prescription for optimizing healthcare outcomes*. Somers, NY: IBM® Institute for Business Value.
- Freund, Y., & Mason, L. (1999). The alternating decision tree learning algorithm. *International Conference on Machine Learning*, 124–133.
- Freund, Y., & Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory* (pp. 23–37). Springer.
- Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, 56, 229–238.
- Hagland, M. (2011). Mastering Readmissions: Laying the Foundation for Change. *Healthcare Informatics*, 28(4), 10–16.
- Haklai, Z., Maron, J., Aborva, M., Gordon, S., Zeltz, I., & Apelboim, Y. (2013). *Admissions to Internal Departments 2000–2012*. Jerusalem: The Ministry of Health.
- Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, 15(6), 1437–1447.
- Hammill, B. G., Curtis, L. H., Fonarow, G. C., Heidenreich, P. A., Yancy, C. W., Peterson, E. D., & Hernandez, A. F. (2011). Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circulation: Cardiovascular Quality and Outcomes*, 4(1), 60–67.
- He, H., Wang, J., Graco, W., & Hawkins, S. (1997). Application of neural networks to detection of medical fraud. *Expert Systems with Applications*, 13(4), 329–336.
- Herbrich, R., Graepel, T., & Campbell, C. (2001). Bayes Point Machines. *Journal of Machine Learning Research*, 1(Aug), 245–279.
- Hon, C. P., Pereira, M., Sushmita, S., Teredesai, A., & De Cock, M. (2016). Risk stratification for hospital readmission of heart failure patients: A machine learning approach. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 491–492). ACM.
- Horn, S. D., Sharkey, P. D., Buckle, J. M., Backofen, J. E., Averill, R. F., & Horn, R. A. (1991). The relationship between severity of illness and hospital length of stay and mortality. *Medical Care*, 29(4), 305–317.
- Kachalia, A., Gandhi, T. K., Puopolo, A. L., Yoon, C., Thomas, E. J., Griffey, R., ... Studdert, D. M. (2007). Missed and delayed diagnoses in the emergency department: A study of closed malpractice claims from 4 liability insurers. *Annals of Emergency Medicine*, 49(2), 196–205.
- Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *JAMA*, 306(15), 1688–1698.
- Kelly, M., Sharp, L., Dwane, F., Kelleher, T., & Comber, H. (2012). Factors predicting hospital length-of-stay and readmission after colorectal resection: A population-based study of elective and emergency admissions. *BMC Health Services Research*, 12(1), 765.
- Khanna, S., Boyle, J., & Good, N. (2014). Precise prediction for managing chronic disease readmissions. In *36th Annual International Conference of the IEEE: Engineering in Medicine and Biology Society (EMBC)* (pp. 2734–2737). IEEE.
- Kirby, S. E., Dennis, S. M., Jayasinghe, U. W., & Harris, M. F. (2010). Patient related factors in frequent readmissions: The influence of condition, access to services and patient choice. *BMC Health Services Research*, 10(1), 309.
- Kociol, R. D., Lopes, R. D., Clare, R., Thomas, L., Mehta, R. H., Kaul, P., ... Armstrong, P. W. (2012). International variation in and factors associated with hospital readmission after myocardial infarction. *JAMA*, 307(1), 66–74.
- Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of Healthcare Information Management*, 19(2), 65.
- Kohli, R., & Tan, S. S.-L. (2016). Electronic health records: How can researchers contribute to transforming healthcare? *MIS Quarterly*, 40(3), 553–573.
- Kossovsky, M. P., Perneger, T. V., Sarasin, F. P., Bolla, F., Borst, F., & Gaspoz, J.-M. (1999). Comparison between planned and unplanned readmissions to a department of internal medicine. *Journal of Clinical Epidemiology*, 52(2), 151–156.
- Krumholz, H. M., Chen, Y.-T., Wang, Y., Vaccarino, V., Radford, M. J., & Horwitz, R. I. (2000). Predictors of readmission among elderly survivors of admission with heart failure. *American Heart Journal*, 139(1), 72–77.
- Leeds, I. L., Sadiraj, V., Cox, J. C., Gao, X. S., Pawlik, T. M., Schnier, K. E., & Sweeney, J. F. (2017). Discharge decision-making after complex surgery: Surgeon behaviors compared to predictive modeling to reduce surgical readmissions. *The American Journal of Surgery*, 213(1), 112–119.
- Lindrooth, R. C., & Weisbrod, B. A. (2007). Do religious nonprofit and for-profit organizations respond differently to financial incentives? The hospice industry. *Journal of Health Economics*, 26(2), 342–357.
- Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 491–502.
- Liu, X., Liu, Y., Lv, Y., Li, C., Cui, Z., & Ma, J. (2015). Prevalence and temporal pattern of hospital readmissions for patients with type I and type II diabetes. *BMJ Open*, 5(11), e007362.

- Marang-van de Mheen, P. J., Lingsma, H. F., Middleton, S., Kievit, J., & Steyerberg, E. W. (2014). Evaluation of quality of care using registry data: The interrelationship between length-of-stay, readmission and mortality and impact on hospital outcomes. *BMJ quality & safety*, 23(4), 350–351.
- Mayer, P. H., Yaron, M., & Lowenstein, S. R. (2010). Learning to use an emergency department information system: Impact on patient length of stay. *Western Journal of Emergency Medicine*, 11(4), 329–332.
- Mesgarpour, M., Chausalet, T., & Chahed, S. (2016). Risk modelling framework for emergency hospital readmission, using hospital episode statistics inpatient data. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 219–224). IEEE.
- Mesgarpour, M., Chausalet, T., & Chahed, S. (2017). Ensemble risk model of emergency admissions (ERMER). *International Journal of Medical Informatics*, 103, 65–77.
- Meydan, C., Haklai, Z., Gordon, B., Mendlovic, J., & Afek, A. (2015). Managing the increasing shortage of acute care hospital beds in Israel. *Journal of Evaluation in Clinical Practice*, 21(1), 79–84.
- Muus, K., Knudson, A., Klug, M., Gokun, J., Sarrazin, M., & Kaboli, P. (2010). Effect of post-discharge follow-up care on re-admissions among us veterans with congestive heart failure: A rural-urban comparison. *Rural Remote Health*, 10(2), 1447.
- Ou, G., & Murphey, Y. L. (2007). Multi-class pattern classification using neural networks. *Pattern Recognition*, 40(1), 4–18.
- Philbin, E. F., Dec, G. W., Jenkins, P. L., & DiSalvo, T. G. (2001). Socioeconomic status as an independent risk factor for hospital readmission for heart failure. *The American Journal of Cardiology*, 87(12), 1367–1371.
- Ramachandran, S., Erraguntla, M., Mayer, R., & Benjamin, P. (2007). Data mining in military health systems-clinical and administrative applications. In *IEEE International Conference on Automation Science and Engineering*, 2007 (pp. 158–163). IEEE.
- Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V., McCoy, T., & Perlis, R. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*, 6(10), e921.
- Schneider, E. B., Hyder, O., Brooke, B. S., Efron, J., Cameron, J. L., Edil, B. H., ... Pawlik, T. M. (2012). Patient readmission and mortality after colorectal surgery for colon cancer: Impact of length of stay relative to other clinical factors. *Journal of the American College of Surgeons*, 214(4), 390–398.
- Shelton, P., Sager, M. A., & Schraeder, C. (2000). The community assessment risk screen (CARS): Identifying elderly persons at risk for hospitalization or emergency department visit. *American Journal of Managed Care*, 6(8), 925–933.
- Silverstein, M. D., Qin, H., Mercer, S. Q., Fong, J., & Haydar, Z. (2008). Risk factors for 30-day hospital readmission in patients? 65 years of age. In *Baylor University Medical Center Proceedings* (p. 363). Baylor University Medical Center.
- Spratt, D. I., Finkelstein, J. S., Butler, J. P., Badger, T. M., & Crowley, Jr., W. F. (1987). Effects of increasing the frequency of low doses of gonadotropin-releasing hormone (GnRH) on gonadotropin secretion in GnRH-deficient men. *The Journal of Clinical Endocrinology & Metabolism*, 64(6), 1179–1186.
- Stiglic, G., Brzan, P. P., Fijacko, N., Wang, F., Delibasic, B., Kalousis, A., & Obradovic, Z. (2015). Comprehensible predictive modeling using regularized logistic regression and comorbidity based features. *PLOS ONE*, 10(12), e0144439.
- Stone, J., & Hoffman, G. J. (2010). *Medicare hospital readmissions: Issues, policy options and ppaca*. Report for Congress.
- Teo, B. K. (2012). *Exafs: Basic principles and data analysis*. New York, NY: Springer Science & Business Media.
- Thayer, C., Bruno, J., & Remorenko, M. B. (2013). Using data analytics to identify revenue at risk: Predictive and comparative analytics have the potential to drive improved value by pinpointing areas where proactive steps can better support optimal revenue cycle performance—as well as the organization's mission. *Healthcare Financial Management*, 67(9), 72–80.
- Tsai, P.-F. J., Chen, P.-C., Chen, Y.-Y., Song, H.-Y., Lin, H.-M., Lin, F.-M., & Huang, Q.-P. (2016). Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. *Journal of Healthcare Engineering*.
- Vest, J. R. (2009). Health Information Exchange and Healthcare Utilization. *Journal of Medical Systems*, 33(3), 223–231.
- van Walraven, C., Dhalla, I. A., Bell, C., Etchells, E., Stiell, I. G., Zarnke, K., ... Forster, A. J. (2010). Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Canadian Medical Association Journal*, 182(6), 551–557.
- Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70, 287–299.
- Welch, H. G., Larson, E. H., Hart, L. G., & Rosenblatt, R. A. (1992). Readmission after surgery in Washington State rural hospitals. *American Journal of Public Health*, 82(3), 407–411.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Amsterdam: Morgan Kaufmann.
- Yang, C., Delcher, C., Shenkman, E., & Ranka, S. (2016). Predicting 30-day all-cause readmissions from hospital inpatient discharge data. In *2016 IEEE 18th International Conference on 2016 IEEE 18th International Conference on* (pp. 1–6). IEEE.
- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.-F., & Hua, L. (2012). Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, 36(4), 2431–2448.
- Zheng, B., Zhang, J., Yoon, S. W., Lam, S. S., Khasawneh, M., & Poranki, S. (2015). Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications*, 42(20), 7110–7120.

Appendix A. Explanation of the terminology

True positive (TP): Correctly identified – The classification model predicts a new patient as a readmission and the patient's actual outcome is “readmission”.

False positive (FP): Incorrectly identified – The classification model predicts a new patient as a readmission and the patient's actual outcome is “non-readmission”.

True negative (TN): Correctly rejected – The classification model predicts a new patient as a non-readmission and the patient's actual outcome is “non-readmission”.

False negative (FN): Incorrectly rejected – The classification model predicts a new patient as a non-readmission and the patient's actual outcome is “readmission”.

Accuracy: The proportion of correct predictions to the total number of predictions. $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$

Precision: Positive Predictive Value or Precision is the proportion of positive cases that were correctly identified by the model. $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$

Recall Sensitivity or Recall is the proportion of actual positive cases that are correctly identified (out of all future readmissions). $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$

F1 Score: The harmonic mean of Precision and Recall. $\text{F1} = 2\text{TP} / (2\text{TP} + \text{FP} + \text{FN})$

Threshold: The cut-off value above which a value belongs to one class and all other values belong to the other class. For example, if the threshold is 0.5, any patient who scored more than or equal to 0.5 is identified as needing readmission, and everyone else belongs to non-readmission class (<http://mund-consulting.com/Blog/understanding-evaluate-model-in-microsoft-azure-machine-learning/>)

Positive/negative label: Binary models predict one of two values. The meaning of these values can be anything, depending on the data; in our case, if a person was readmitted. Statisticians and data scientists use the terms “positive” and “negative”. The decision to assign “positive” must refer to the value that drives the decision (<https://blogs.msdn.microsoft.com/andreasderuiter/2015/02/09/performance-measures-in-azure-ml-accuracy-precision-recall-and-f1-score/>).

ROC: the receiver operating characteristic curve illustrates the performance of a binary classifier system as its discrimination threshold varies. An ROC plot examines the trade-off between the two types of positive signals in the models. The horizontal axis represents the ratio of false positives, and the vertical axis shows the ratio of true positives.

AUC: Area under the ROC curve (the value is between 0 and 1) measures the area under the curve plotted with true positives on the y axis and false positives on the x axis. A diagonal line indicates a poor model for predictions, hence the AUC, for large data-sets, should never be lower than 0.5.

Appendix B. List of Abbreviations

BDT	Boosted decision tree
BPM	Bayes point machine
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
EHR	Electronic health record
ED	Emergency department also known as an accident & emergency department (A&E) in UK, emergency room (ER) in US and worldwide, emergency ward (EW) in US and worldwide

ENT	Ears, nose, and throat
GUI	Graphical user interface
HIE	Health information exchange
HIT	Health information technology
HMO	Health maintenance organisation
IS	Information systems
IT	Information technology
LOS	Length of stay
ML	Machine learning
OR	Odds ratio
ROC	Receiver operating characteristic
SVM	Support vector machine
UTI	Urinary tract infection

Appendix C. Profiling patients that were predicted very good vs. those that were predicted very poorly

In this appendix, the comparison was made on the validation data-set that included 30% of all patients (2536 patients) because the remaining 70% of the data-set was used for training the model.

The results show that for both classifiers, the variables differed significantly for the two categories according to the following attributes: Age, Days Between Referrals, the previous number of ED visits and the number of prior Readmissions (only for the BDT). Additionally, the following diagnoses differed significantly: Heart Diseases (only for the BDT), COPD and Enteritis and Colitis.

Hence, we suggest that patients that are correctly predicted for readmissions and for non-readmissions are younger, have more days between the current referral and the previous one, have higher numbers of previous ED visits, have lower numbers of previous readmissions, have more diagnoses of heart diseases, less diagnoses COPD, more diagnoses of Enteritis And Colitis. The summary tables are shown below.

Table C1. BDT classifier.

Data characteristics	TP + TN <i>n</i> = 2163 (85.3%)	FP + FN <i>n</i> = 373 (14.7%)
Age*	66.1 ± 19.5	68.6 ± 18.2
Insurance (main HMO %)	89%	87%
Creatinine Result	1.49 ± 1.5	1.55 ± 1.4
EHR Use	82%	83%
Length of Stay	2.55 ± 4.3	2.92 ± 5.7
DaysBetweenReferrals*	104.8 ± 130.1	90.53 ± 125.5
Prior#Visits**	7.51 ± 1.56	7.27 ± 1.4
PriorReAdmission-30Day_Sum**	2.47 ± 1.7	2.65 ± 1.2
HeartDisease**	33%	26%
NephritisUTI	21%	21%
COPD*	9%	14%
BloodFormingOrgan	10%	10%
Pneumonia	10%	10%
EnteritisAndColitis*	9%	6%
CerebrovascularDisease	10%	8%

Notes: Data are the mean (±SD) or number of subjects (proportion). The four groups were compared for significant differences according to the types of variables, using T-Tests, proportion tests, etc.

Table C2. Logistic regression classifier.

Data characteristics	TP + TN $n = 2536$ (81.6%)	FP + FN $n = 467$ (18.4%)
Age**	66 ± 19.7	68.7 ± 18
Insurance (main HMO %)	89%	87%
Creatinine Result	1.48 ± 1.5	1.55 ± 1.6
EHR Use	82%	84%
Length of Stay	2.56 ± 4.5	2.81 ± 4.6
DaysBetweenReferrals+	104.74 ± 131	93.6 ± 122.3
Prior#Visits*	7.51 ± 1.6	7.32 ± 1.4
PriorReAdmission30Day_Sum	2.48 ± 1.7	2.55 ± 1.2
HeartDisease	32%	30%
NephritisUTI	21%	21%
COPD+	9%	13%
BloodFormingOrgan	10%	10%
Pneumonia	12%	10%
EnteritisAndColitis*	9%	7%
CerebrovascularDisease	10%	8%