Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30085642

Original Paper

Forecasting Daily Patient Outflow From a Ward Having No Real-Time Clinical Data

Shivapratap Gopakumar, MTech; Truyen Tran, PhD; Wei Luo, PhD; Dinh Phung, PhD; Svetha Venkatesh, PhD

Centre for Pattern Recognition and Data Analytics, Deakin University, Geelong Waurn Ponds, Australia

Corresponding Author:
Shivapratap Gopakumar, MTech
Centre for Pattern Recognition and Data Analytics
Deakin University
Building KA
75 Pigdons Road
Geelong Waurn Ponds, 3216
Australia
Phone: 61 3 5227 1266
Fax: 61 3 5227 2028
Email: shivapratap@gmail.com

Abstract

Background: Modeling patient flow is crucial in understanding resource demand and prioritization. We study patient outflow from an open ward in an Australian hospital, where currently bed allocation is carried out by a manager relying on past experiences and looking at demand. Automatic methods that provide a reasonable estimate of total next-day discharges can aid in efficient bed management. The challenges in building such methods lie in dealing with large amounts of discharge noise introduced by the nonlinear nature of hospital procedures, and the nonavailability of real-time clinical information in wards.

Objective: Our study investigates different models to forecast the total number of next-day discharges from an open ward having no real-time clinical data.

Methods: We compared 5 popular regression algorithms to model total next-day discharges: (1) autoregressive integrated moving average (ARIMA), (2) the autoregressive moving average with exogenous variables (ARMAX), (3) k-nearest neighbor regression, (4) random forest regression, and (5) support vector regression. Although the autoregressive integrated moving average model relied on past 3-month discharges, nearest neighbor forecasting used median of similar discharges in the past in estimating next-day discharge. In addition, the ARMAX model used the day of the week and number of patients currently in ward as exogenous variables. For the random forest and support vector regression models, we designed a predictor set of 20 patient features and 88 ward-level features.

Results: Our data consisted of 12,141 patient visits over 1826 days. Forecasting quality was measured using mean forecast error, mean absolute error, symmetric mean absolute percentage error, and root mean square error. When compared with a moving average prediction model, all 5 models demonstrated superior performance with the random forests achieving 22.7% improvement in mean absolute error, for all days in the year 2014.

Conclusions: In the absence of clinical information, our study recommends using patient-level and ward-level data in predicting next-day discharges. Random forest and support vector regression models are able to use all available features from such data, resulting in superior performance over traditional autoregressive methods. An intelligent estimate of available beds in wards plays a crucial role in relieving access block in emergency departments.

(JMIR Med Inform 2016;4(3):e25) doi:10.2196/medinform.5650

KEYWORDS

patient flow; discharge planning; predictive models

Introduction

Demand for health care services has become unsustainable [1,2]. This is largely due to increase in population and life expectancy, escalating costs, increased patient expectations, and workforce issues [3]. Despite increased demands, the number of inpatient beds in hospitals has come down by 2% since the last decade

[2,4]. Efficient bed management is key to meeting this rising demand and reducing health care costs.

Daily discharge rate can be a potential real-time indicator of operational efficiency [5]. From a ward-level perspective, a good estimate of next-day discharges will enable hospital staff to foresee potential problems such as changes in number of available beds and changes in number of required staff. Efficient forecasting reduces bed crisis and improves resource allocation. This foresight can help accelerate discharge preparation, which has huge cost on clinical staff and educating patients and family, requiring postdischarge planning [6,7]. However, studying patient flow from general wards offers several challenges.

Ward-level discharges incorporate far greater hospital dynamics that are often nonlinear [8]. Accessing real-time clinical information in wards can be difficult because of administrative and procedural barriers, such data may not be available for predictive applications. Because the diagnosis coding is performed after discharge, there is little information about medical condition or variation in care quality in real time. In addition, factors other than patient condition play a role in discharge decisions [5,9,10].

The current practice of bed allocation in general wards of most hospitals involves a hospital staff/team, who use past information and experience, to schedule and assign beds [11]. Modern machine learning techniques can be used to aid such decisions and help understand the underlying process. As an example, Figure 1 illustrates a decision tree trained on past discharges and ward occupancy statistics, which models the daily discharge pattern from an open ward in a regional Australian hospital. Although the absence of patient medical information affected forecast performance, the decision rules provide important insight into the discharge process.

Motivated by this result, we address the open problem of forecasting daily discharges from a ward with no real-time clinical data. Specifically, we compare the forecasting performance of 5 popular regression models: (1) the classical autoregressive integrated moving average (ARIMA), (2) the autoregressive moving average with exogenous variables (ARMAX), (3) k-nearest neighbor (kNN) regression, (4) random forest (RF) regression, and (v) support vector regression (SVR). Our experiments were conducted on commonly available data from a recovery ward (heath wing 5) in Barwon Health, a regional hospital in Victoria, Australia. The ARIMA and kNN models are built from daily discharges from ward. To account for the seasonal nature of discharges, the ARMAX model included day of the week and ward occupancy statistics. We identified and constructed 20 ward-level and 88 patient-level predictors to derive the RF and SVR models.

Forecasting accuracy was measured using 3 metrics on a held out set of 2511 patient visits in the year 2014. When compared with a naive forecasting method of using the mean of last week discharges, we demonstrate through our experiments that (1) using regression methods for forecasting discharge outperforms naive forecasting, (2) SVR and RF models outperform the autoregressive methods and kNN, (3) an RF model derived from 108 features has the minimum error for next-day forecasts.

The significance of our study is in identifying the importance of foreseeing available beds in wards, which could help relieve emergency access block [12].

Patient length of stay directly contributes to hospital costs and resource allocation. Long-term forecasting in health care aims to model bed and staffing needs over a period of months to years. Cote and Tucker categorize the common methods in health care demand forecasting as percent adjustment, 12-month moving average, trendline, and seasonalized forecast [13]. Although each of these methods is built from historical demand, seasonalized forecasting provides more realistic results as it takes into account the seasonal variations and trends in the data. Mackay and Lee [3] advise modeling the patient flow in health care institutions for tactical and strategic forecasting. To this end, compartmental modeling [14,15], queuing models [16,17] and simulation models [17-20] have been applied to analyze patient flow. To understand long-term patient flow, studies analyze metrics such as bed occupancy [3,8,14,19,21,22], patient arrivals [23], and individual patient length of stay [19,24-27].

On the other hand, our work implements short-term forecasting. The short-term forecasting methods are concerned with hourly and daily forecasts from a single unit in a care environment. The most popular unit of interest is the emergency or acute care department because this is often a key performance indicator metric in assessing quality of care [28,29].

Figure 1. Decision tree modeling of total discharges from an open ward from day of the week and ward occupancy (previous day occupation) data for 5 years. The leaves represent total number of patient discharges.
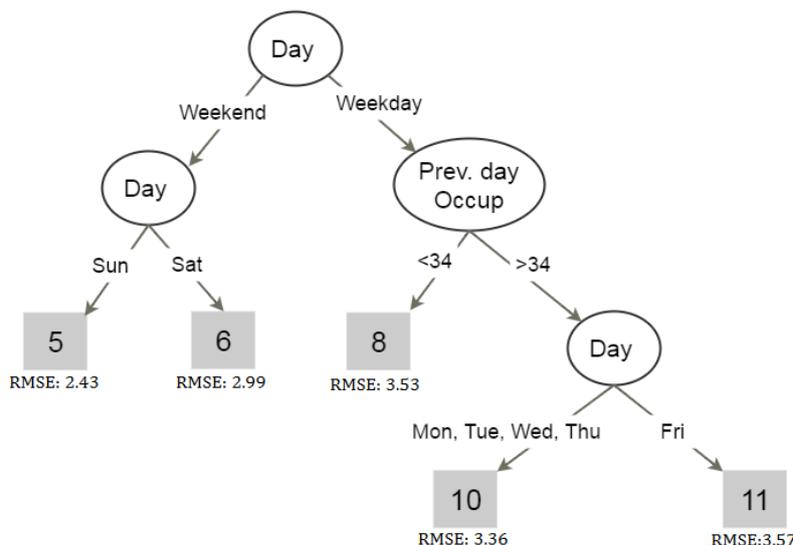
Time Series and Smoothing Methods

When looking at discharges as time series, autoregressive moving average models are the most popular [30-32]. Exponential smoothing techniques have also been used to forecast monthly [33] and daily patient flow [34].

Jones et al used the classical ARIMA to forecast daily bed occupancy in emergency department of a European hospital [30]. The model which included seasonality terms demonstrated reasonable performance to predict bed occupancy. The authors speculated whether nonlinear forecasting techniques could improve over ARIMA. A recent study confirmed the effectiveness of this forecasting technique in a US hospital setting [35]. ARIMA models were also successfully used to forecast the number of occupied beds during a SARS outbreak in a Singapore hospital [36]. A recent study used patient attendances in a pediatric emergency department to model daily demand using ARIMA [37].

Jones et al [34] compared the ARIMA mode with exponential smoothing and artificial neural networks to forecast daily patient volumes in emergency department. The study revealed no single model to be superior and concluded that seasonal patterns play a major role in daily demand.

Simulation Methods

Modeling using simulation is typically used to study the behavior of complex systems. An early work in investigated the effects of emergency admissions on daily bed requirements in acute care, using discrete-event stochastic simulation modeling [38]. Sinreich and Marmor [39] proposed a guide for building a simulation tool based on data from emergency departments of 5 Israeli hospitals. Their method analyzed the flow of patients clustered into 8 types along with time elements. The simulation demonstrated that patient processes are better characterized by type of the patients, rather than specific hospitals visited. Yeh and Lin used a simulation model to characterize patient flow through a hospital emergency department and reduced waiting times using a genetic algorithm [40]. A similar experiment was carried out in a geriatric department using a combination of discrete event simulation and queuing model to analyze bed occupancy [19].

Regression for Forecasting

Regression models analyze the relationship between the forecasted variable and features in the data. Linear regression that encoded monthly variations was used to forecast patient admissions over a 6-month horizon and outperformed quadratic and autoregressive models [41]. Another study used clustering and Principle Component Analysis PCA to find significant predictors from patient data to model emergency length of stay using linear regression [42]. A nonlinear approach using regression trees was proposed in forecasting patient admissions which demonstrated superior performance over a neural net framework [43].

Barnes et al used 10 predictors to model real-time inpatient length of stay in a 36-bed unit using an RF model [24].

Nonlinear regression is better suited to model the changing dynamics of patient flow. To characterize the outflow of patients from the ward, we resort to regression using RF, kNN, and SVR. In the area of pattern recognition, kNNs [44] are the most effective method that exploits repeated patterns. The kNN algorithm has been successfully applied to forecast to histogram time series in financial data [45]. The nonparametric regression using kNN has been successfully demonstrated for short-term traffic forecasting [46,47] and electricity load forecasting [48,49]. However, kNN regression has not been studied for patient flow.

Another powerful and popular regression technique, SVR, uses kernel functions to map features into a higher dimensional space to perform linear regression. Though this technique has not seen much application in medical forecasting, support vector machines have been successful in financial market prediction, electricity forecasting, business forecasting, and reliability forecasting [50].

Apart from the standard autoregressive methods, we use kNN, RFs, and SVR in forecasting next-day discharges. Because

discharge patterns repeat over time, kNN regression can be applied to search for a matching pattern from past discharges. RFs and SVR regression are powerful modelling techniques requiring minimum tuning to effectively handle nonlinearity in the hospital processes.

Recently, RF forecasting was used to predict total patient discharges from a 36 bed unit in an urban hospital [24]. Apart from 4 demographic and 2 timing predictors, this study used 3 clinical predictors for patients: (1) reason for visit: identified by a physician and recorded using International Classification of Diseases: version 9 (ICD-9) diagnosis codes [51], (2) observation status: assigned to patients for monitoring purpose, and (3) pending discharge location. Total number of discharges was estimated from aggregate of individual patient length of stay.

The absence of real-time clinical information in our data makes calculating patient length of stay impossible. Instead, we resort to modelling next-day discharges by observing previous discharge patterns and examining demographics and flow characteristics in the ward.

Methods

Data

Our study used retrospective data collected from a recovery ward in Barwon Health, a large public health provider in Victoria, Australia serving about 350,000 residents. Ethics approval was obtained from the Hospital and Research Ethics Committee at Barwon Health (number 12/83) and Deakin University. The total number of available beds depended on the number of staff assigned to the ward. On average, the ward had 36 staffed beds, but fluctuated between 20 and 80 beds with varying patient flow. The physicians in the ward had no teaching responsibilities.

Table 1. Tables in hospital database used in our data collection.

Tables	Columns
Patients	1. Patient ID
	2. Age
	3. Gender
Ward Stay	1. Admission ID
	2. Name of the ward
	3. Time (entry, exit)
	4. Bed ID
Admissions	1. Patient ID
	2. Admission ID
	3. Time (admit, discharge)
	4. Patient Class (21 categories)
	5. Admission type (7 categories)

Table 2. Cohort details.

Cohort	Stats
Total patient visits	12,141
Unique patients	10,610
Length of stay (mean, median, IQR[a])	4.26, 3, 5
Discharges per day (mean, median, IQR)	8.7, 8, 5
Admissions per day (mean, median, IQR)	8.6, 8, 5
Mean ward occupancy, IQR	30.9, 4
Gender	54.8% Female
Age (mean, median)	66, 63.23

[a]IQR, interquartile range.

The data for our study came from three tables in the hospital database, as shown in Table 1. Additional real-time data that described patient condition or disease progression were unavailable because diagnosis coding using medical codes is done after discharge. Patient flow was collected for a period of 4 years. Using the admission and discharge times for each patient, we calculated the daily discharges from our ward in study. A total of 12,141 patients were admitted into the ward with a median discharge of 8 patients per day from January 1,

2010, to December 31, 2014. Table 2 summarizes the main characteristics of our data.

A time series decomposition of our data revealed strong seasonal variations and high nonlinearity in daily discharge patterns. There was a defined weekly pattern–discharge from ward peaked on Fridays and dropped significantly on weekends (see Figure 2). This seasonal nature is in tune with previous studies [9,32]. Aggregating the daily discharges into a monthly time series revealed defined monthly patterns (see Figure 3). The data displayed no significant trend. In addition, the daily discharge

pattern was found to be highly nonlinear. Our forecasting methods must be able to handle such data dynamics.

We describe the following diverse methods that are applicable to forecasting under complex data dynamics: (1) ARIMA, (2) autoregressive moving, (3) forecasting using kNN discharge patterns, (4) RF, and (5) SVR. Autoregressive methods model the temporal linear correlation between nearby data points in the time series. Nearest patterns lift this linearity assumption and assumes that short periods form repeated patterns. Finally, RF and SVR look for a nonlinear functional relationship between the future outcomes and descriptors in the past.

Figure 2. Mean admissions and discharges per day from ward.
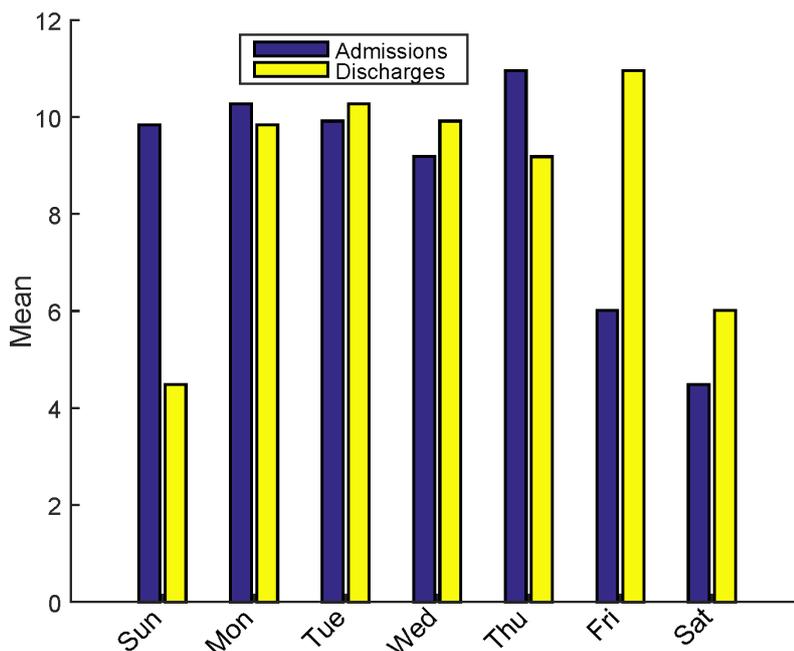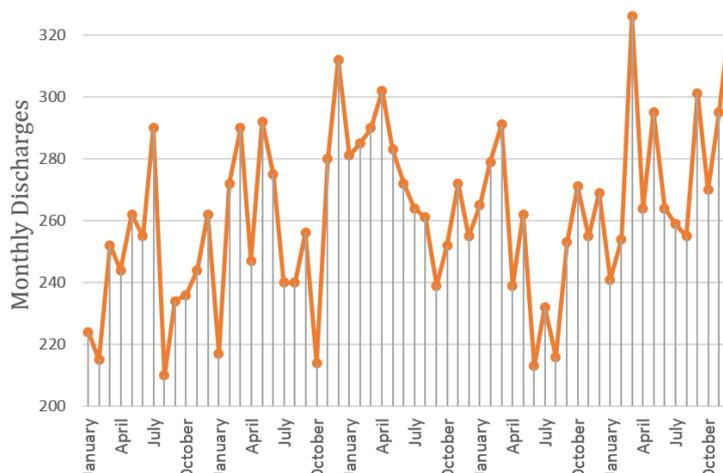

Figure 3. Time series of monthly discharges from ward.

Forecasting Methods

Autoregressive Integrated Moving Average

Time-series forecasting methods can analyze the pattern of past discharges and formulate a forecasting model from underlying temporal relationships [52]. Such models can then be used to

extrapolate the discharge time series into the future. ARIMA models are widely used in time-series forecasting. Their popularity can be attributed to ease of model formulation and interpretability [53]. ARIMA models look for linear relationships in the discharge sequence to detect local trends and seasonality. However, such relationships can change over

time. ARIMA models are able to capture these changes and update themselves accordingly. This is done by combining autoregressive (AR) and moving average (MA) models. Autoregressive models formulate discharge at time t=y_t, as a linear combination of previous discharges. On the other hand, moving averages models characterize as linear combination of previous forecast errors. For ARIMA model, the discharge time series is made stationary using differencing. Let $\emptyset$ be autoregressive parameters, θ be moving average parameters, and ϵ be the forecast errors. Such an ARIMA model can be defined as shown in Figure 4, where μ is a constant. By varying p and q, we can generate different models to fit the data. Box Jenkins method [54] provides a well-defined approach for model identification and parameter estimation. In our work, we choose the auto.arima() function from the forecast package [55] in R [56] to automatically select the best model.

Figure 4. Classical ARIMA model.

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_q \epsilon_{t-i}$$

Autoregressive Moving Average With Exogenous Variables (ARMAX)

Dynamic regression techniques allow adding additional explanatory variables, like day of the week and number of current patients in the ward, to autoregressive models. The autoregressive moving ARMAX modifies ARIMA model by including depending external variable x_t at time t, as shown in Figure 5. We model x_t using features from the hospital database.

Figure 5. ARIMA model with exogenous variable xt.

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_q \epsilon_{t-i} + \beta x_t$$
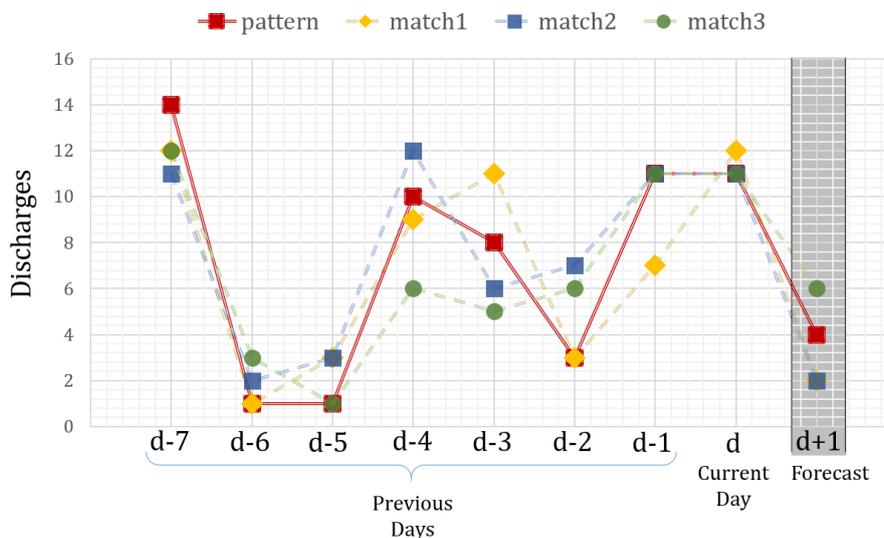
Detecting Discharge Patterns Using k-Nearest Neighbors

The kNN algorithm takes advantage of the locality in data space. We assume that the next-day discharge depends on the discharges happening in previous days. Using kNN principles, we can do a regression to forecast the next-day discharge. Let y_d represent number of discharges on the current day: d. To forecast the next day discharge: y_{d+1}, we look at the discharges over the past p days as: disch_vec=$[y_{d-p}: y_d]$. Using Euclidean distance metric, we find k closest matches to disch_vec from the training data. An estimate of next-day discharge: $\hat{y}_{d+1}$, is calculated as a measure of the next-day discharges of the k matched patterns: $(y_{match})_i \in (1:k)$. Figure 6 shows an example of kNN based forecasting. Here, disch_vec in red $[y_{d-7}: y_d]$ results in 3 matches from the training data. For simplicity, we have plotted the matched patterns alongside disch_vec, although they had occurred in the past. The next-day forecast $\hat{y}_{d+1}$ becomes a measure of $(y_{match})_i$, where $(y_{match})_i \in (1:3)$ is the $(d +1)^{th}$ term of each of the matched patterns [57].

One popular method of calculating $\hat{y}_{d+1}$ is by minimizing the weighted quadratic loss (Figure 7), where w_i takes values between 0 and 1, with $\sum_{i=1}^{k} w_i = 1$. However, there are 2 main drawbacks making it less desirable for our data. First, the quadratic loss is sensitive to outliers. Second, a robust estimate of $\{ w_i \}$ becomes difficult.

Our data contain significant noise, causing large variations in next-day forecasts of the k matched patterns. We illustrate this problem in Figure 8. For a given day, kNN regression returns 125 matched patterns. The next-day forecasts from each k=125 patterns displayed significant variations. In such scenario, we resort to estimating $\hat{y}_{t+1}$ by minimizing the robust loss (Figure 9).

Figure 6. k-nearest neighbor forecasting example with k=3 and P=7.

XSL·FO
RenderX

Figure 7. Calculating ŷd+1 by minimizing the weighted quadratic loss.

$$\hat{y}_{d+1} = \min_y \sum_{i=1}^{k} w_i \left((y_{match})_i - y\right)^2$$

$$= \sum_{i=1}^{k} w_i (y_{match})_i$$

Figure 8. Scatterplot of next-day forecast using k-nearest neighbor for a given day. X-axis represents each matched nearest-neighbor pattern. Y-axis represents the next day forecast of that matched pattern.
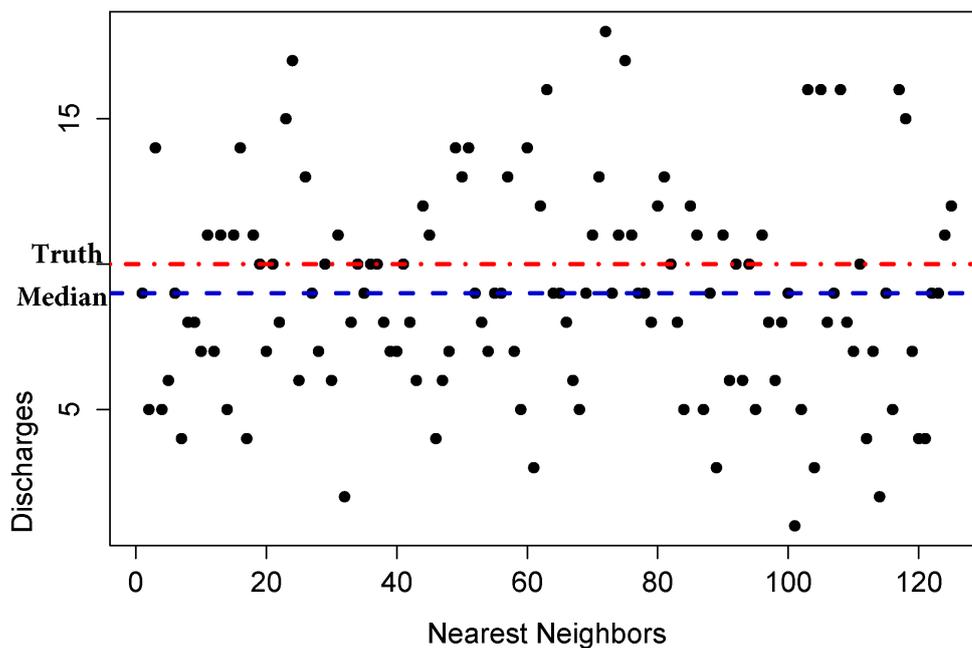

Figure 9. Estimating ŷt+1 by minimizing the robust loss.

$$\hat{y}_{t+1} = \min_y \sum_{i=1}^{k} |(y_{match})_i - y|$$

$$= \text{median}(y_{match})_{i=1\ to\ k}$$

Random Forest

In this approach, we assume the next-day discharge as a function of historical descriptor vector: x. We use each day in the past as a data point, where the next-day discharge is the outcome, and the short period before the discharge are used to derive descriptors. The RF used in this paper is currently one of the most powerful methods to model the function $y = f(x)$ [58,59]. An RF is an ensemble of regression trees. A regression tree approximates a function $f(x)$ by recursively partitioning the descriptor space. At each region R_p, the function is approximated as shown in Figure 10, where $|R_p|$ is the number of data point falling in region R_p. The RF creates a diverse collection of random trees by varying the subsets of data points to train the trees and the subsets of descriptors at each step of space partitioning. The final outcome of RF is an average of all trees in the ensemble. Since tree growing is a highly adaptive process, it can discover any nonlinear function to any degree of approximation if given enough training data. However, the flexibility makes regression tree prone to overfitting, that is, the inability to generalize to unseen data. This requires controlling the growth by setting the number of descriptors per partitioning step, and the minimum size of region R_p.

The voting leads to great benefits: reduce the variations per tree. The randomness helps combat against overfitting. There is no assumption about the distribution of data or the form of the function (x). There is controllable quality of fits.

Figure 10. Random forests formulation of next day discharges (y) from historical descriptors (x).

$$f(x) = \frac{1}{|R_p|} \sum_{x_j \in R_p} y_j$$

Support Vector Regression

The historical descriptor vector *x,* used in the RF model can also be used to build a SVR model [60]. Given the set of data $\{(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)\}$, where each $x_i \epsilon R^m$ denotes the input descriptor for the corresponding next day forecast $y_i \epsilon R^1$, a regression function takes the form: $\hat{y}_i = f(x_i)$. SVR works by (1) mapping the input space of x_i into a higher dimensional space using a nonlinear mapping function: φ, (2) performing a linear regression in this higher dimensional space. In general, we can express the regression function as: $f(x) = (w\varphi(x)) + b,$ where, $w \epsilon R^m$ is the weights and $b \epsilon R^1$ is the bias term. Vapnik [60] proposed the ∈-insensitive loss function for SVR, which takes the form as shown in Equation 1 in Figure 11. The loss function L_ϵ tolerates errors that are smaller than the threshold: $\epsilon,$ resulting in a "tube" around the true discharge values. Model parameters can be estimated by minimizing the cost function as shown in Equation 2 in Figure 11, where *C* is a constant that penalizes error in training data.

In our work, we use an RBF kernel [61] for mapping our input data to higher dimensional feature space. RBF kernels are a good choice for fitting our nonlinear discharge pattern because of its ability to map the training data to an infinite dimensional

space and easy implementation. The solution to the dual formulation of SVR cost function is detailed in [60,62].

Figure 11. The SVR learning model.

$$L_\epsilon(f(\boldsymbol{x}) - \boldsymbol{y}) = \begin{cases} |f(\boldsymbol{x}) - \boldsymbol{y}|, & |f(\boldsymbol{x}) - \boldsymbol{y}| \geq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

Equation 1. ∈-insensitive loss function.

$$R = C \times \frac{1}{n} L_\epsilon(f(\boldsymbol{x}) - \boldsymbol{y}) + \frac{1}{2} \|w\|^2$$

Equation 2. SVR cost function.

Experiments

We extracted all data from the database tables (as in Table 1) for our ward in study. Patient flow was analyzed for a period of 5 years. We formatted our data as a matrix where each row corresponds to a day and each column represents a feature (descriptor). Two main groups of features were identified: (1) ward level and (2) patient level. Our feature creation process resulted in 20 ward-level and 88 patient-level predictors, as listed in Table 3. The ward-level descriptor: trend of next-day discharge was calculated by fitting a locally weighted polynomial regression [63] from past discharges. An example of this regression fitting is shown in Figure 12.

Table 3. Features constructed from ward data in hospital database.[a]

Type	Predictor	Description
Ward-level	Seasonality	Current day-of-week, current month
	Trend	Calculated using locally weighted polynomial regression from past discharges on the same weekday
	Admissions	Number of admissions during past 7 days
	Discharges	Number of discharges during past 7 days, number of discharges in previous 14th day and 21st day
	Occupancy	Ward occupancy in previous day
Patient-level	Admission type	5 categories
	Patient referral	49 categories
	Patient class	21 categories
	Age category	8 categories
	Number of wards visited	4 categories
	Elapsed length of stay	Calculated daily for each patient in the ward

[a] The random forest and support vector regression models used the full set of features. The ARMAX (autoregressive moving average with exogenous variables) model used seasonality and occupancy. All other models were derived from daily discharges.

Figure 12. An example of the discharge trend, as derived from a locally weighted polynomial regression model.
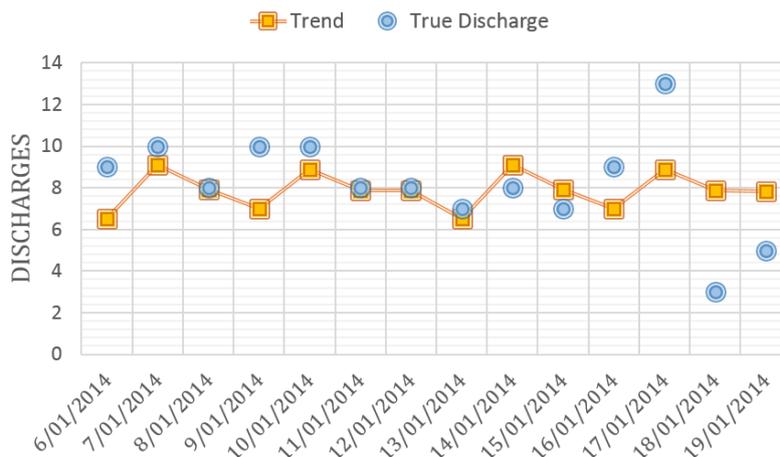
Evaluation Protocol

Our training and testing sets are separated by time. This strategy reflects the common practice of training the model using data in the past and applying it on future data. Training data consisted of 1460 days from January 1, 2010, to December 31, 2013. Testing data consisted of 365 days in the year 2014. The characteristics of the training and validation cohort are shown in Table 4. Most stays were short, around 65% of patients stayed for less than 5 days.

Table 4. Characteristics of training and validation cohorts.

Categorization	Training (2010-2013)	Testing (2014)
Total days	1460	365
Mean discharges per day	8.47	9.17
Number of admissions	9630	2511
Gender		
Male	4329 (44.9%)	1135 (45.2%)
Female	5301 (55.1%)	1376 (54.8%)
Mean age (years)	63.65	61.62
Length of stay		
1-4 days	6377 (66.22%)	1636 (65.15%)
5 or more days	3253 (33.78%)	875 (34.85%)

Baseline Forecasting

The current hospital strategy involves using past experience to foresee available beds. To compare the efficiency of our proposed approaches, we model the following baselines: (1) Naive forecasting using the last day of week discharge: since our data were found to have defined weekly patterns, we model the next day discharge as the number of discharges for the same day during previous week; (2) naive forecasting using mean of last week discharges: to better model the variation and noise in weekly discharges, we model the next-day discharge as the mean of discharges during previous 7 days; and (3) naive forecasting using mean of last 3-week discharges: to account for the monthly and weekly variations in our data, we use mean of daily discharges over the past 3 weeks to model the next-day discharge.

Measuring Forecast Performance

We compare the next-day forecasts of our proposed approaches with the baseline methods on the measures of mean forecast error, mean absolute error, symmetric mean absolute percentage error, and root mean square error [64,65]. If y_t is the measured discharge at time t, f_t is the forecasted dishcharge at time t, we can define the following:

• Mean forecast error (MFE): is used to gauge model bias and is calculated as MFE = mean$(y_t - f_t)$

• For an ideal model, MFE = 0. If MFE > 0, the model tends to underforecast. When MFE < 0, the model tends to overforecast.

• Mean absolute error (MAE): is the average of unsigned errors: MAE = mean$| y_t - f_t |$.

MAE indicates the absolute size of the errors.

• Root mean square error (RMSE) is a measure of the deviation of forecast errors. It is calculated as: RMSE = $\sqrt{\text{mean}(y_t - f_t)^2}$

Due to squaring and averaging, large errors tend to have more influence over RMSE. In contrast, individual errors are weighted

equally in MAE. There has been much debate on the choice of MAE or RMSE as an indicator of model performance [66,67].

•Symmetric mean absolute percentage error (sMAPE): It is scale independent and hence can be used to compare forecast performance between different data series. It overcomes 2 disadvantages of mean absolute percentage error (MAPE) namely, (1) the inability to calculate error when the true discharge is zero and (2) heavier penalties for positive errors than negative errors. sMAPE is a more robust estimate of forecast error and is calculated as: $sMAPE = mean(200[|y_t - f_t|/y_t + f_t])$. However, sMAPE ranges from −200% to 200%, giving it an ambiguous interpretation [68].

Results

Model Performance

In this section, we describe the results of comparing our different forecasting methods. The model parameters for kNN forecast, RF, and SVR models were tuned to minimize forecast errors.

For kNN regression, the optimum value of pattern length: d and number of nearest neighbours: k, was obtained by analyzing forecast RMSE for values $d \in (1,100)$ and $k \in (5,1000)$. Minimum RMSE of 3.77 was obtained at $d=70$ and $k=125$.

The SVR parameters C (penalty cost) and ϵ (amount of allowed error) were determined by choosing the best value from a grid search, that minimized the model RMSE. Similarly, the optimum number of variables in building each node of the RF was chosen by examining its effect on minimizing the out-of-bag estimate.

We compared the naive forecasting methods with our proposed 5 models using MFE, MAE, RMSE, and sMAPE. The results are summarized in Table 5, whereas Figure 13 compares the distribution of actual discharges with different model forecasts.

Table 5. Forecast accuracy of different models.

Model	Mean forecast error	Mean absolute error	Symmetric mean absolute percentage error	Root mean square error	Mean absolute error improve over naïve
Naive forecast					
Using discharge from last weekday	0.03	3.81	45.70 %	4.95	
Using mean of last week discharges	0.02	3.57	41.68 %	4.42	
Using mean of last 3-week discharges	0.04	3.44	40.14%	4.34	
ARIMA[a]	0.06	3.27	38.32 %	4.15	4.9 %
ARMAX[b]	-0.01	2.99	34.86 %	3.84	13.1 %
k-nearest neighbor	1.09	2.88	34.92 %	3.77	16.3 %
Support vector regression	0.73	2.75	32.88%	3.64	20.1 %
Random forest	0.44	2.66	31.86 %	3.49	22.7 %

[a] ARIMA: autoregressive integrated moving average

[b] ARMAX: autoregressive moving average with exogenous variables

The naive forecasts are unable to capture all variations in the data and resulted in the maximum error when compared with other models.

The variations in seasonality and trend are better captured in ARIMA and ARMAX models. The time series consisting of past 3-month discharges were used to generate the next-day discharge forecast. The ARMAX model also included the day of week and ward occupancy as exogenous variables, which resulted in better forecast performance over ARIMA.

Interestingly, kNN was more successful than ARIMA and ARMAX in capturing the variations in discharge, demonstrating about 3% improvement in MAE, when compared with ARMAX. However, the kNN model tends to under forecast (MFE = 1.09),

possibly because of resorting to median values for forecast. In comparison, RF and SVR forecast models demonstrated better performance. This can be expected because they are derived from all the 108 features. However, RF demonstrated a relative improvement of 3.3 % in MAE over SVR model (see Table 5). When looking at forecast errors for each day of week, RF model confirmed better performance, as shown in Figure 14.

The process of SVR with RBF kernel maps all data into a higher dimensional space. Hence, the original features responsible for forecast cannot be recovered, and the model acts as a black box. Alternatively, RF algorithm returns an estimate of importance for each variable for regression. Examining the features with high importance could give us a better understanding of the discharge process.

Figure 13. Comparison of actual and forecasted discharges from ward for each day in 2014.
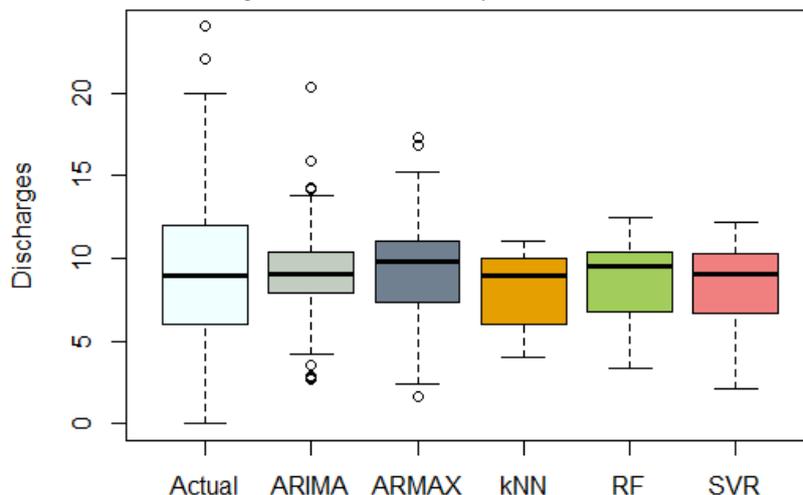
Figure 14. Forecast error in predicting each day of week in 2014.
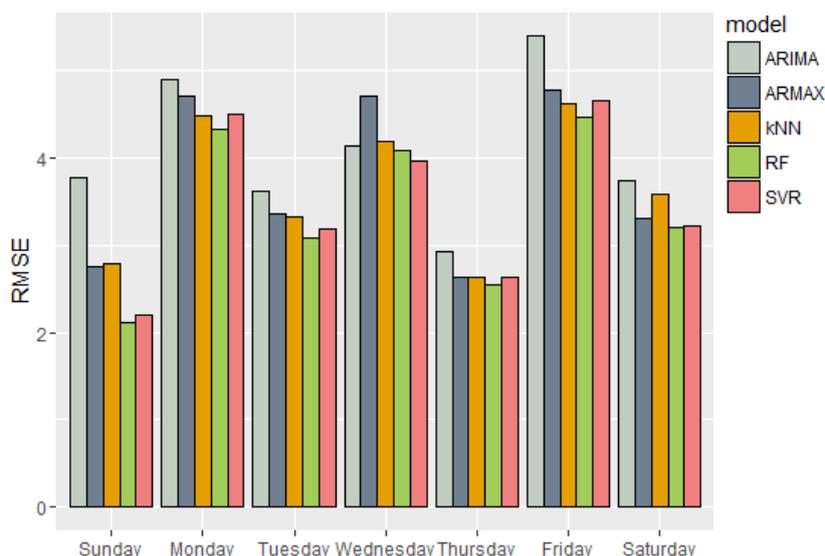
Feature Importance in the Random Forest model

The features in random forecast model were ranked on importance scores. The top 10 significant features are described as follows. The day of week for the forecast proved to be the most important feature. Other features were number of patients in the ward during the day of forecast, the trend of discharges measured using locally weighted polynomial regression, number of discharges in past 14th day, number of discharges in past 21st day, number of patients who had visited only one previous ward, the number of males in the ward, number of patients labelled as: "public standard," and current month of forecast.

Discussion

Principal Findings

Improved patient flow and efficient bed management is key to counter escalating service and economic pressures in hospitals. Predicting next-day discharges is crucial but has been seldom studied for general wards. When compared with emergency and acute care wards, predicting next-day discharges from a general ward is more challenging because of the nonavailability of real-time clinical information. The daily discharge pattern is seasonal and irregular. This could be attributed to management of hospital processes such as ward rounds, inpatient tests, and medication. The nonlinear nature of these processes contributes to unpredictable length of stay even in patients with similar diagnosis.

Typically, for open wards, a floor manager uses previous experience to foresee the number of available beds. In this paper, we attempt to model total number of next-day discharges using 5 methods. We have compared the forecasting performance using MAE, RMSE, and sMAPE. Our predictors are extracted from commonly available data in the hospital database. Although the kNN method is simple to implement, requiring no special expertise, software packages for other models are available for all common platforms. These models can be implemented by the analytics staff in hospital IT department and can be easily integrated into existing health information systems.

In our experiments, forecast based on RF model outperformed all other models. Forecasting error rate is 31.9% (as measured by sMAPE) which is in the same ballpark as the recent work of [24], though we had no real-time clinical information. An

RF model makes minimum assumptions about the underlying data. Hence, it is the most flexible, and at the same time, comes with great overfitting control. Similarly, SVR also demonstrated superior performance, compared with the autoregressive and kNN models. The RBF kernel maps the features into a higher dimensional space during the regression process. Hence, the physical meaning of the features is lost, making it difficult to interpret the model. Finally, RFs and SVR are able to handle more features. This extra information in the form of patient demographics and past admission and discharge statistics contributed to improve the predictive performance when compared with other models.

The kNN regression also performed well as it assumes only the locality in the data. But it is not adaptive, and thus less flexible in capturing complex patterns. The kNN regression assumes similar patterns in past discharges extrapolate to similar future discharge, which is not true for daily discharges from ward. ARMAX model outperformed the traditional ARIMA forecasts since it incorporated seasonal information as external regressors. As expected, a naive forecast of using the median of past discharges performed worst.

We noticed a weekly pattern (Figure 2) and monthly pattern (Figure 3) in discharges from the ward. Other studies have also confirmed that discharges peak on Friday and drop during weekends [5,9,10]. This "weekend effect" could be attributed to shortages in staffing or reduced availability of services like sophisticated tests and procedures [10,69]. This suggests discharges are heavily influenced by administrative reasons and staffing.

Feature importance score from an RF model helps in identifying the features contributing to the regression process. The day of forecast proved to be one of the most important features in the RF model. Other important features included trend based on nonlinear regression of past weekdays, number of discharges in the past days, ward occupancy in previous day, number of males in the ward, and number of general patients in ward.

When looking at for each day of the week, the RF and SVR model consistently outperformed other models. Sundays and Thursdays proved to be the easiest to predict for all models (Figure 14). This can be expected since these days had the least variation in our data. Fridays proved to be the most difficult to forecast. Retraining the RF model by omitting "day of the week" increased the forecast error by 1.39% (as measured by sMAPE).

Patient length of stay is inherently variable, partly due to the complex nonlinear structure of medical care [8]. The number of discharges from a ward is strongly related to the length of stay of the current patients in the ward. Hence, the variability in ward-level discharges is compounded by the variability in individual patient length of stay. In our study, the daily discharge pattern from ward shows great variation for each day of week. Apart from patient level details, we believe that a knowledge of hospital policies is also required to capture such nonlinearity.

Practical Significance

In our study, we were able to validate that the weekend patterns affect discharges from a general ward. The RF model was able to give a reasonable estimate of number of next-day discharges from the ward. Clinical staff can use this information as an aid to decisions regarding staffing and resource utilization. This foresight can also aid discharge planning such as communication and patient transfer between wards or between hospitals.

An estimate of number of free beds can also help reduce emergency department (ED) boarding time and improve patient flow [12,23]. ED boarding time is the time spent by a patient in emergency care when a bed is not available in the ward. ED boarding time severely reduces the hospital efficiency. High bed occupancy in ward directly contributes to ED overcrowding [70]. In our data, 42.81% of patients were admitted from the emergency care. An estimate of daily forecasts can be helpful in deciding the number of beds in wards to ease patient flow.

Study Limitations

We acknowledge the following limitations in our study. First, we focused only on a single ward. However, it was a ward with different patient types, and hence the results could be an indication for all general wards. Second, we did not use patient clinical data to model discharges. This was because clinical diagnosis data were available only for 42.81% of patients who came from emergency. In a general ward, clinical coding is not done in real time. However, we believe that incorporating clinical information to model patient length of stay could improve forecasting performance. Third, we did not compare our forecasts with clinicians/managing nurses. Finally, our study is retrospective. However, we have selected prediction period separated from development period. This has eliminated possible leakage and optimism.

Conclusion

This study set out to model patient outflow from an open ward with no real-time clinical information. We have demonstrated that using patient-level and ward-level features in modelling forecasts outperforms the traditional autoregressive methods. Our proposed models are built from commonly available data and hence could be easily extended to other wards. By supplementing patient-level clinical information when available, we believe that the forecasting accuracy of our models can be further improved.

XSL•FO

RenderX

Conflicts of Interest

None declared.

References

1. Kalache A, Gatti A. Active ageing: a policy framework. Adv Gerontol 2003;11:7-18. [Medline: 12820516]
2. OECD. A Disease-based Comparison of Health Systems: What is Best and at what Cost?. Paris: OECD publishing; 2003.
3. Mackay M, Lee M. Choice of models for the analysis and forecasting of hospital beds. Health Care Manag Sci 2005 Aug;8(3):221-230. [Medline: 16134435]
4. Alijani A, Hanna GB, Ziyaie D, Burns SL, Campbell KL, McMurdo ME, et al. Instrument for objective assessment of appropriateness of surgical bed occupancy: validation study. BMJ 2003 Jun 7;326(7401):1243-1244 [FREE Full text] [doi: 10.1136/bmj.326.7401.1243] [Medline: 12791738]
5. Wong H, Wu RC, Caesar M, Abrams H, Morra D. Real-time operational feedback: daily discharge rate as a novel hospital efficiency metric. Qual Saf Health Care 2010 Dec;19(6):e32. [doi: 10.1136/qshc.2010.040832] [Medline: 20724394]
6. Connolly M, Deaton C, Dodd M, Grimshaw J, Hulme T, Everitt S, et al. Discharge preparation: do healthcare professionals differ in their opinions? J Interprof Care 2010 Nov;24(6):633-643. [doi: 10.3109/13561820903418614] [Medline: 20919958]
7. Connolly M, Grimshaw J, Dodd M, Cawthorne J, Hulme T, Everitt S, et al. Systems and people under pressure: the discharge process in an acute hospital. J Clin Nurs 2009 Feb;18(4):549-558. [doi: 10.1111/j.1365-2702.2008.02551.x] [Medline: 19192004]
8. Harper P, Shahani AK. Modelling for the Planning and Management of Bed Capacities in Hospitals. The Journal of the Operational Research Society 2002;53(1):11-18 [FREE Full text] [WebCite Cache]
9. Wong H, Wu RC, Tomlinson G, Caesar M, Abrams H, Carter MW, et al. How much do operational processes affect hospital inpatient discharge rates? J Public Health (Oxf) 2009 Dec;31(4):546-553 [FREE Full text] [doi: 10.1093/pubmed/fdp044] [Medline: 19465455]
10. van WC, Bell CM. Risk of death or readmission among people discharged from hospital on Fridays. CMAJ 2002 Jun 25;166(13):1672-1673 [FREE Full text] [Medline: 12126321]
11. Daniels MJ, Kuhl ME, Hager E. Forecasting Hospital Bed Availability Using Simulation and Neural Networks. In: Proceedings of IIE Annual Conference. 2005 Presented at: IIE Annual Conference and Exposition; May 14-18, 2005; Atlanta.
12. Luo W, Cao J, Gallagher M, Wiles J. Estimating the intensity of ward admission and its effect on emergency department access block. Stat Med 2013 Jul 10;32(15):2681-2694. [doi: 10.1002/sim.5684] [Medline: 23172783]
13. Côté MJ, Tucker SL. Four methodologies to improve healthcare demand forecasting. Healthc Financ Manage 2001 May;55(5):54-58. [Medline: 11351811]
14. McClean S, Millard PH. A decision support system for bed-occupancy management and planning hospitals. IMA J Math Appl Med Biol 1995;12(3-4):249-257. [Medline: 8919561]
15. McClean S, Millard PH. A three compartment model of the patient flows in a geriatric department: a decision support approach. Health Care Manag Sci 1998 Oct;1(2):159-163. [Medline: 10916595]
16. el-Darzi E, Vasilakis C, Chaussalet T, Millard PH. A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department. Health Care Manag Sci 1998 Oct;1(2):143-149. [Medline: 10916593]
17. Mills TM. A mathematician goes to hospital. Australian Mathematical Society Gazette 2004;31(5):320-327.
18. Costa A, Ridley SA, Shahani AK, Harper PR, De SV, Nielsen MS. Mathematical modelling and simulation for planning critical care capacity. Anaesthesia 2003 Apr;58(4):320-327 [FREE Full text] [Medline: 12648112]
19. el-Darzi E, Vasilakis C, Chaussalet T, Millard PH. A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department. Health Care Manag Sci 1998 Oct;1(2):143-149. [Medline: 10916593]
20. Hoot N, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting emergency department crowding: a discrete event simulation. Ann Emerg Med 2008 Aug;52(2):116-125. [doi: 10.1016/j.annemergmed.2007.12.011] [Medline: 18387699]
21. Mackay M. Practical experience with bed occupancy management and planning systems: an Australian view. Health Care Manag Sci 2001 Feb;4(1):47-56. [Medline: 11315885]
22. Gorunescu F, McClean SI, Millard PH. Using a queueing model to help plan bed allocation in a department of geriatric medicine. Health Care Manag Sci 2002 Nov;5(4):307-312. [Medline: 12437280]
23. Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. Predicting emergency department inpatient admissions to improve same-day patient flow. Acad Emerg Med 2012 Sep;19(9):E1045-E1054 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01435.x] [Medline: 22978731]
24. Barnes S, Hamrock E, Toerper M, Siddiqui S, Levin S. Real-time prediction of inpatient length of stay for discharge prioritization. J Am Med Inform Assoc 2016 Apr;23(e1):e2-e10. [doi: 10.1093/jamia/ocv106] [Medline: 26253131]

25. Levin SR, Harley ET, Fackler JC, Lehmann CU, Custer JW, France D, et al. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. Crit Care Med 2012 Nov;40(11):3058-3064. [doi: 10.1097/CCM.0b013e31825bc399] [Medline: 22824935]

26. Clark DE, Ryan LM. Concurrent prediction of hospital mortality and length of stay from risk factors on admission. Health Serv Res 2002 Jun;37(3):631-645 [FREE Full text] [Medline: 12132598]

27. Marshall A, Vasilakis C, El-Darzi E. Length of stay-based patient flow models: recent developments and future directions. Health Care Manag Sci 2005 Aug;8(3):213-220. [Medline: 16134434]

28. Kulinskaya E, Kornbrot D, Gao H. Length of stay as a performance indicator: robust statistical methodology. IMA Journal of Management Mathematics 2005;16(4):369-381.

29. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a modified-delphi approach. Acad Emerg Med 2002 Nov;9(11):1131-1139 [FREE Full text] [Medline: 12414461]

30. Jones SA, Joy MP, Pearson J. Forecasting demand of emergency care. Health Care Manag Sci 2002 Nov;5(4):297-305. [Medline: 12437279]

31. Littig SJ, Isken MW. Short term hospital occupancy prediction. Health Care Manag Sci 2007 Feb;10(1):47-66. [Medline: 17323654]

32. Lin RC, Pasupathy KS, Sir MY. Estimating Admissions and Discharges for Planning Purposes-Case of an Academic Health System. Advances in Business and Management Forecasting 2011;8:115-128.

33. Lin WT. Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches. International Journal of Forecasting 1989 Jan;5(2):195-208. [doi: 10.1016/0169-2070(89)90087-3]

34. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. Acad Emerg Med 2008 Feb;15(2):159-170 [FREE Full text] [doi: 10.1111/j.1553-2712.2007.00032.x] [Medline: 18275446]

35. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. Acad Emerg Med 2009 Apr;16(4):301-308 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00356.x] [Medline: 19210488]

36. Earnest A, Chen MI, Ng D, Sin LY. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Serv Res 2005;5:36 [FREE Full text] [doi: 10.1186/1472-6963-5-36] [Medline: 15885149]

37. Kadri F, Harrou F, Chaabane S, Tahon C. Time series modelling and forecasting of emergency department overcrowding. J Med Syst 2014 Sep;38(9):107. [doi: 10.1007/s10916-014-0107-0] [Medline: 25053208]

38. Bagust A, Place M, Posnett JW. Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. BMJ 1999 Jul 17;319(7203):155-158 [FREE Full text] [Medline: 10406748]

39. Sinreich D, Marmor Y. Emergency department operations: the basis for developing a simulation tool. IIE transactions 2005;37(3):233-245.

40. Yeh JY, Lin WS. Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. Expert Systems with Applications 2007;32(4):1073-1083.

41. Boyle J, Wallis M, Jessup M, Crilly J, Lind J, Miller P, et al. Regression forecasting of patient admission data. Conf Proc IEEE Eng Med Biol Soc 2008;2008:3819-3822. [doi: 10.1109/IEMBS.2008.4650041] [Medline: 19163544]

42. Combes C, Kadri F, Chaabane S. Predicting Hospital Length Of Stay Using Regression Models: Application To Emergency Department. 2014 Presented at: 10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM; November 5-7, 2014; France p. 14.

43. Garcia KA, Chan PK. Estimating Hospital Admissions with a Randomized Regression Approach. 2012 Presented at: 11th International Conference on Machine Learning and Applications ICMLA 2012; December 12-15, 2012; Boca Raton, FL p. 179-184.

44. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory 1967;13(1):21-27.

45. Arroyo J, Mat'e C. Forecasting histogram time series with k-nearest neighbours methods. International Journal of Forecasting 2009;25(1):192-207.

46. Davis GA, Nihan NL. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. Journal of Transportation Engineering 1991;117:178-188.

47. Zhang L, Liu Q, Yang W, Wei N, Dong D. An improved k-nearest neighbor model for short-term traffic flow prediction. Procedia-Social and Behavioral Sciences 2013;96:653-662.

48. Al-Qahtani FH, Crone SF. Multivariate k-nearest neighbour regression for time series data—A novel algorithm for forecasting UK electricity demand. 2013 Presented at: Neural Networks (IJCNN), The 2013 International Joint Conference; August 4-9, 2013; Dallas, TX. [doi: 10.1109/IJCNN.2013.6706742]

49. Tsakoumis AC, Vladov SS, Mladenov VM. Daily load forecasting based on previous day load. 2002 Presented at: Neural Network Applications in Electrical Engineering, 2002 (NEUREL'02); May 12-17, 2002; Honolulu p. 83-86. [doi: 10.1109/NEUREL.2002.1057973]

XSL·FO

RenderX

50.   Sapankevych N, Sankar R. Time series prediction using support vector machines: a survey. IEEE Computational Intelligence Magazine 2009;4(2):24-38.

51.   Centers for Disease Control and Prevention. 1978. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) URL: http://www.cdc.gov/nchs/icd/icd9cm.htm [accessed 2016-07-12] [WebCite Cache ID 6iwJ90Tnk]

52.   Chatfield C. The analysis of time series: an introduction. In: The Analysis of Time Series: An Introduction, Sixth Edition. London: Chapman & Hall/CRC; 2003.

53.   Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics 2014;15:276 [FREE Full text] [doi: 10.1186/1471-2105-15-276] [Medline: 25123979]

54.   Box G, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. Englewood Cliffs, NJ: Prentice Hall; 1994.

55.   Hyndman R, Khandakar JY. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 2008;26(3):1-22.

56.   R Development Core Team. GBIF. Vienna, Austria; 2011. R: A Language and Environment for Statistical Computing URL: http://www.R-project.org/ [accessed 2016-07-12] [WebCite Cache ID 6iwep4Khx]

57.   Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 1992;46(3):175-185.

58.   Breiman L. Random forests. Machine learning 2001;45(1):5-32.

59.   Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2001.

60.   Vapnik V. The nature of statistical learning theory. New York: Springer; 2000.

61.   Schölkopf B, Tsuda K, Vert JP. Kernel methods in computational biology. Cambridge, MA: MIT Press; 2004.

62.   Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and computing 2004;14(3):199-222.

63.   Cleveland W, Grosse E, Shyun W, Chambers J, Hastie T. Local regression models. In: Statistical models in S. New York: Chapman and Hall; 1992:309-376.

64.   Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting 2006 Oct;22(4):679-688. [doi: 10.1016/j.ijforecast.2006.03.001]

65.   Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev V. A survey of forecast error measures. World Applied Sciences Journal 2013;24:171-176.

66.   Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 2005;30(1):79-82.

67.   Chai T, Draxler R. Root mean square error (RMSE) or mean absolute error (MAE)? Geoscientific Model Development Discussions 2014;7:1525-1534.

68.   Hyndman RJ. Another look at forecast-accuracy metrics for intermittent demand. Foresight: The International Journal of Applied Forecasting 2006;4(4):43-46.

69.   Lee LH, Swensen SJ, Gorman CA, Moore RR, Wood DL. Optimizing weekend availability for sophisticated tests and procedures in a large hospital. Am J Manag Care 2005 Sep;11(9):553-558 [FREE Full text] [Medline: 16159045]

70.   Forster AJ, Stiell I, Wells G, Lee AJ, van WC. The effect of hospital occupancy on emergency department length of stay and patient disposition. Acad Emerg Med 2003 Feb;10(2):127-133 [FREE Full text] [Medline: 12574009]

Abbreviations

ARIMA: autoregressive intensive moving average
ARMAX: autoregressive moving average with exogenous variables
ED: emergency department
kNN: k-nearest neighbor
MAE: mean absolute error
MAPE: mean absolute percentage error
MFE: mean forecast error
RF: random forest
RMSE: root mean square error
sMAPE: symmetric mean absolute percentage error
SVR: support vector regression

XSL·FO

RenderX