Smart Health 18 (2020) 100145

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

GAM feature selection to discover predominant factors for mortality of weekend and weekday admission to the ICUs

Yi Cai, Jianian Zheng, Xiaoliang Zhang, Haotian Jiang, Ming-Chun Huang *

Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, 44106, USA

ARTICLE INFO

Keywords:
ICU
GAM feature Selection
Weekend
Weekday

ABSTRACT

"Weekend effect" on mortality is a common controversial topic among many hospitals. Although the mortality of patients associated with weekend Intensive Care Units (ICUs) admission has been demonstrated slightly higher than that of patients admitted on weekdays in many studies, the underlying causal mechanisms and the potential factors are not clear at present. In this study, we extract medical record features from the database and propose a Generalized Additive Model (GAM) feature selection method to identify the main contribution features for analyzing this issue. The GAM feature selection system could rank candidate features by its importance, which turns out to be effective in reducing the complexity of a medical issue. The best lists of features are acquired by the weekday GAM model and the weekend GAM model separately. Fourteen out of forty-one features are identified for the reduced list of features. Both models' reduced lists of features have ten identical characteristics, and the other four are different. The prediction accuracy with the reduced list of features is 79.90% for the weekday model and 78.56% for the weekend model. The contrast experiment has validated the feature ranking results. Furthermore, variables of the same feature classes are also different from weekday admission to weekend admission. We expect that the proposed GAM feature selection method could contribute to solving more medical issues in the future.

1. Introduction

Several studies have found that mortality among patients admitted on weekends was higher than for those patients admitted on weekdays (Freemantle et al., 2015; Walker et al., 2017). This kind of phenomenon exists in the Intensive Care Unit (ICU) outstandingly. Many people purely consider that hospitals provide routine care on weekdays and hospital staffing deficiency on weekends. However, it is much convincing to find the predominant factors of this problem from the perspective of medical data representation.

ICU as a special department is widely existing in many hospitals for critically sick patients who are with severe and life-threatening illnesses or injuries. It can continuously provide intensive treatment medicine, advanced medical resources and equipment, and highly trained doctors or nurses who are specialized in caring for critical patients. Generally, patients who are admitted to ICU are usually in severe conditions. Barnett et al. have proposed that patients admitted to an ICU on the weekend usually have a higher risk for death and more ICU length of stay by analyzing the dataset of 156,136 patients admitted to 38 ICUs in 28 hospitals (Barnett et al., 2002). With the question to the study, Allen et al. wanted to determine whether weekend admission to the ICU increases the risk of dying in

* Corresponding author.
E-mail addresses: yxc757@case.edu (Y. Cai), jxz852@case.edu (J. Zheng), xxz585@case.edu (X. Zhang), hxj172@case.edu (H. Jiang), ming-chun.huang@case.edu (M.-C. Huang).

https://doi.org/10.1016/j.smhl.2020.100145
Received 27 November 2019; Received in revised form 31 October 2020; Accepted 3 November 2020
Available online 11 November 2020
2352-6483/

the hospital with a database of a total of 29,084 patients admitted to medical, surgical, and multispecialty ICUs, which share relatively uniform staffing and availability of diagnostic and therapeutic options (Ensminger et al., 2004). Typically, the dataset used in this paper is Medical Information Mart for Intensive Care III (MIMIC-III) (Johnson et al., 2016), which has the information of patients admitted to 7 ICUs of one hospital.

Although the results of excess mortality associated with weekend admission have been demonstrated in many studies based on the various datasets, the underlying causal mechanisms and the potential factors are not clear at present. Fortunately, the prevalence of electronic health records offers an opportunity to extract patients' clinically relevant survival data, vital signs, medications, laboratory measurements, etc (Celi et al., 2014). Moreover, a proper regressing model is also essential for retrieving the corresponding features. Comprehensive analysis is quite important for looking into solutions in the future as well. To address the unanswered problems, we conducted a study to find out what are the potential predominant factors leading to mortality of weekday and weekend admission to the ICU. A feature selection framework that used GAM to find the important features was proposed. To evaluate the feature selection framework, we apply the machine learning technique, Bagged tree, on the application of weekdays or weekends death prediction. By comparing the two models' feature selection results, four attributes transfer, temperature, number of services accepted and DRG_mortality were picked out for further studying the features of weekdays and weekends admission. Moreover, variables of the same features have been proved relevant to the day of admission (weekdays or weekends) as well in the prediction models contrast experiments. One attribute (admission_type) is selected to demonstrate this distinction through statistical analysis. The aims of this paper can be concluded as:

- *Analyzing if there is a mortality difference for patients admitted to the intensive care unit on a weekend and those admitted on a weekday based on the MIMIC-III database.*
- *Design a GAM feature selection method.*
- *Using the GAM feature selection method to find out the contribution of each feature and rank them according to their significance.*
- *Analyzing the causes leading to the difference in mortalities with prediction models*

2. Related work

RezaSadeghi et al. (Sadeghi et al., 2018) had proposed using vital heart signals of patients within the first hour of ICU admission to do early hospital mortality prediction based on the MIMIC-III database. Each signal is described in terms of 12 statistical and signal-based features. Their experiment results indicated that heart rate signals can be used for predicting mortality in patients in the care units especially coronary care units (CCUs). However, the proposed method is only applicable to CCU patients who must have the heart signals information. With the MIMIC III medical ICU dataset, Jacob et al. (Calvert et al., 2016) proposed the AutoTriage algorithm with eight common clinical variables to do medical ICU mortality prediction instead of considering the difference between weekday and weekend admission. Furthermore, previous studies suggested that patients are more likely to die in the hospital if they are admitted on a weekend than on a weekday. Researchers from different countries have been dedicated to studying this problem for decades by analyzing different kinds of health-related datasets in different periods. Barba et al. (Barba et al., 2006) have analyzed the clinical data of 35,993 adults (>14 years) patients admitted to the emergency department of Fundación Hospital Alcorcón in Madrid, Spain, and concluded that the risk of mortality within the first 48 h is higher for patients admitted on weekends than for patients admitted on a weekday. Bell et al. (Bell & Redelmeier, 2001) have compared in-hospital mortality among patients admitted on a weekend with that among patients admitted on a weekday by analyzing the Canadian data of all acute care admissions from emergency departments in Ontario. James et al. (James et al., 2010) focused on the investigation of whether patients admitted on a weekend with acute kidney injury (AKI) were more likely to die than those admitted on a weekday with a large database of admissions to acute care, nonfederal hospitals in the United States. NFreemantle et al. (Freemantle et al., 2012) from Britain have tried to assess whether weekend admissions to hospital and/or already being an inpatient on weekend days were associated with any additional mortality risk by developing survivorship models to analyze the medical information of admissions to the English National Health Service (NHS) during the financial year 2009/10. However, there are fewer studies to determine if specific medical information is valuable in studying the weekend effect. Meanwhile, it is worth to summary more analysis that weekend admission to the ICU increases the risk of dying in the hospital from the aspect of extracting features from medical records. Rachel et al. (Meacock et al., 2017) proposed that the reduced availability of care services may result in increasing the admission threshold at weekends, and they also pointed out that a high proportion of the sicker population of patients admitted on weekends combined with this selection effect is partly responsible for the weekend effect. Typically, the most acceptable potential mechanisms for the weekend effect are: (1) insufficient hospital staffing on weekends leading to many restrictions on maintaining the hospital's normal operation; (2) staffing by relatively less experienced staff could hardly ensure accurate services for the patient; (3) worse conditions of the weekends' admissions because of their greater severity. Although these potential mechanisms are proposed in many research studies, it remains unclear which mechanism truly explains the weekend effect (Ranji et al., Sarkar). In this study, we seek to propose a method by finding the contribution of documented medical information of each admitted patient to analyzing the weekend effect. We hope with more detailed medical records included, the dominant features as potential evidence could support medical service providers on determining causal factors and proposing specific solutions.

3. Feature selection and evaluation framework

In Fig. 1, we design a framework to do feature selection and evaluation for this project. The framework mainly contains three parts:

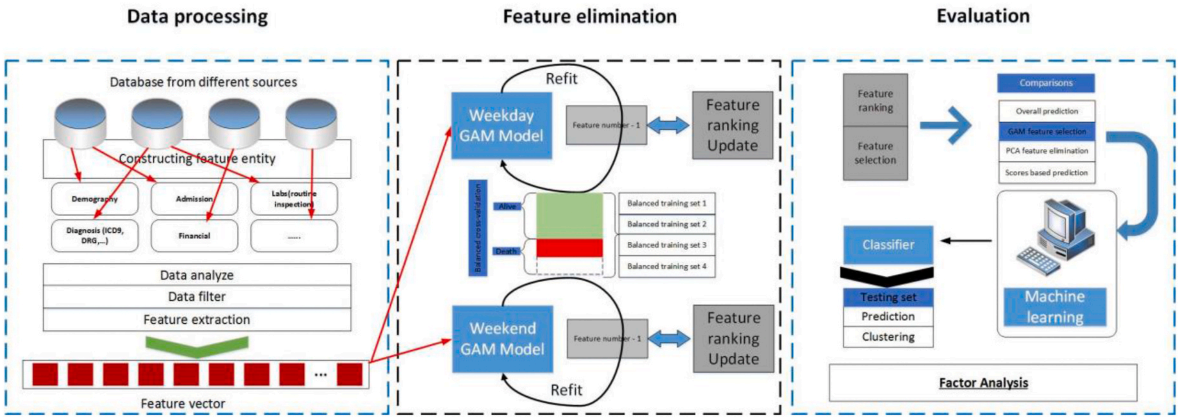

Fig. 1. Feature selection and evaluation framework: 1) Data processing; 2) Feature elimination; 3) Evaluation. The first part is about how to do the data processing and extract corresponding features for analysis according to the MIMIC-III database. The second part is about using GAM model with a recursive feature elimination algorithm to do feature selection. Weekday GAM model and Weekend GAM model are generated separately to find the predominant factors on mortality for each situation. The last part is to evaluate the feature selection results based on classification results with machine learning technique.

1) Data processing; 2) Feature elimination; 3) Evaluation. The first part is about how to do the data processing and extract corresponding features for analysis according to the MIMIC-III database. The second part is about using the GAM model with a recursive feature elimination algorithm to complete the feature selection. Weekday GAM model and Weekend GAM model are generated separately to find the predominant factors on mortality for each situation. The last part is to evaluate the feature selection results based on the contrast experiments and outcome predictions with machine learning techniques.

3.1. Generalized Additive Model (GAM)

GAM is an upgraded version of the Generalized linear model (GLM) which holds the thought that there is a linear relationship between dependent and independent variables (Clark). It can capture non-linear, linear, and non-monotonic relationships between predictor variables and the response, but retain much of the intelligibility of linear models (Lou et al., 2012, pp. 150–158). The difference between GAMs and GLMs is that, in general, linear predictor in GAMs is not forced to be linear, but is the sum of smooth functions of the predictor variables, which can be in many types of forms, represented as $f_i(x)$:

$$g(y) = b_0 + f_1(x_1) + f_2(x_2) + \ldots\ldots + f_P(x_P) \tag{1}$$

In equation (1), the function $g(.)$ is a *link function* that is in the logit model to describe a GAM. The $f_i(x)$, $i = 1, \ldots, p$, are smooth functions that use a scatter plot which is more flexible than parametric models to reveal the relationship between response and predictor (Lou et al., 2012, pp. 150–158). By using GAMs, we can easily get information about the contribution of each x_i.

Implement GAMs in R: Like the t-test in GLMs, F-test (Wikipedia. F-test — Wikip, 2020) is used to find out if the variances between the two populations are significantly different, where an "F Test" is a catch-all term for any test that uses the F-distribution. And F-value, the outcome of this test, contains the information of whether a single variable is significant or a group of variables are jointly significant.

The *"R-sq. (adj)"*, means "adjusted R square" (Miles, 2014), which is a modified version of R-squared that has been adjusted for the number of predictors in the model. If any new terms improve the model more than expectation, the *R-sq. (adj)* will increase, vice versa. Equation (2) shows how to calculate it.

$$R^2_{adj} = 1 - \frac{(n-1)^*(1-R^2)}{n-p-1} \tag{2}$$

Where n is the number of points in the data sample, and p is the number of variables in the model. R^2 shows how well the data points fit into a model. If more new variables have been put into a model, the R^2 will increase, but R^2_{adj} only increases if the added variables are useful. It means that the R^2_{adj} will decrease if the useless variables are added into the model.

In statistics, deviance is a goodness-of-fit statistic for a statistical model. The deviance D of a model is defined through equation 3

$$D = 2^*[l(\beta_{\max}) - l(\beta)]^*\phi \tag{3}$$

where $l(\hat{\beta}_{max})$ is the maximized likelihood of the saturated model and $l(\hat{\beta})$ is the maximized likelihood of the model that has fitted. Deviance explained is just representing the above as the proportion of the total deviance explained by the current model, greater values indicated better fitting models. The GCV score is the minimized generalized cross-validation (GCV) score of the GAM fitted. GCV is used for smoothness selection, smaller values indicated better fitting models.

3.2. Feature selection algorithm

To get performance estimates that incorporate the selection bias due to feature selection, we implement a feature selection algorithm in R. For completeness, pseudo-code in Algorithms 1 is included, which has a more complete definition, the algorithm fits the model with all predictors. Each predictor is ranked by its importance to the model. Let K be a sequence of ordered numbers which are candidate values for the number of predictors to retain ($K_1 > K_2, \ldots$). At each iteration of feature selection, a 4-fold cross-validation strategy is used. In this way, the whole dataset is

arbitrarily divided into four disjoint sections, a balanced training set, three folds shape a balanced training set and the remaining one is utilized to test. The K_i top important predictors are retained by combining the four folds ranking results with the majority, the model is refit, and performance is assessed. The value of K_i with the best performance is determined and the top K_i predictors are used to fit the final model. In the initialization part, we do the data pre-processing and build the first model with all the features. In the iteration part, we do the feature selection. Then multiple lists of the 'best' predictors are generated at each iteration. At the end of the algorithm, a consensus ranking can be used to determine the best predictors to retain. Overall, this kind of method can reduce the effect of selection bias and keep the interpretability.

Algorithm 1 Recursive Feature Elimination

Input: N elements (x_i, y_i), where $y_i = 0, 1$ when do a classification problem, $y_i \in$ R. Here $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ is a feature vector with K features
Output: The final model with weight in front of each predictor. At each iteration of feature selection, the K_i top important predictors are retained, the model is refit, and performance is assessed.
for Each Resampling **do**
/* Initialization*/
Step 1: Pre-process the whole data set (balanced four-fold cross-validation).
Step 2: In each fold, calculate feature importance or rankings and model performances with all the K_{i-1} features.
/* Iteration, for each subset size K_i, i = 1; 2 … */
Step 3: Use majority to get K_i most important predictors by combining the four folds ranking results.
Step 4: Refit the model using K_i and continue the iteration.
end for
Step 5: Determine the final feature importance or ranking.
Step 6: Four-fold cross-validation to partition data into four training sets and four test sets.
Step 7: Use balanced training sets to train classifiers with machine learning methods, such as Bagged tree. Step 8: Calculate the classifier performances using test sets.
Step 9: Evaluate the feature selection outcome based on the average classifier performance.

3.3. Bagged tree

The decision tree is a method commonly used in machine learning and data mining (Lioret al., 2014). When constructing decision trees, the algorithm needs to recursively choose the optimal input feature in each node that best splits the set of subjects, and stop when all subjects at a node have the same target value, or when splitting no longer improve the prediction result. In order to reduce the variance and improve the accuracy of the decision tree, we also use a powerful ensemble method called Bagging (Bootstrap Aggregation) (Breiman, 1996). The bagging method can iteratively generate new training sets by sampling from the original training set uniformly and with replacement. Then the new training sets will be used to construct different decision tree classifiers. Finally, the target label of each subject is the most frequent class predicted by the classifiers.

4. Data processing

4.1. Data resource

MIMIC-III (Medical Information Mart for Intensive Care III) is a large, freely-available database comprising identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The database includes information such as demographics, vital sign measurements made at the bedside (1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital). Well-trained abstractors are responsible for collecting the demographic and clinical data on each patient at the time of admission, transfer, or discharge from the hospital, such as age, gender, source of admission (categorized as elective, newborn, urgent, or emergency). The dataset includes the information of patients admitted to 7 units, which are coronary care unit, cardiac surgery recovery unit, medical intensive care unit, neonatal intensive care unit, neonatal intensive care unit, neonatal ward, surgical intensive care unit, and trauma/surgical intensive care unit.

4.2. Data preparation

The admissions table gives information regarding a patient's admission to the hospital, which contains demographic information like ethnicity, religion, marital status, and specific timestamp of admission, discharge, and death. Medical histories are usually

Table 1
Codebook of selected features for Weekday GAM model and Weekend GAM model.

Variable	Definition	Type	Scale
DRG_CODE	The code represents the Diagnosis-related group (DRG) (The drgcodes 2020), which is a system which classifies hospital cases according to certain groups, which gathers claim information based on ICD diagnoses, procedures, age, sex, discharge status and the presence of complications or comorbidities (Wikipedia. Diagnosis-rela, 2020).	Character	1–9550
DRG_MORTALITY	Provide additional granularity to DRG codes in the All Patients Refined Diagnosis Related Groups (APR-DRGs) type [1]. This subclass is risk of mortality which presents the likelihood of dying [11].	Numeric	0-4 (0:None; 1:Minor; 2:Moderate; 3: Major; 4:Extreme)
DRG_SEVERITY	Provide additional granularity to DRG codes in the APR-DRGs type (The drgcodes , 2020).This subclass is severity of illness, which presents the extent of physiologic decomposition or organ system loss of function (Fontaine, 2004).	Numeric	0-4 (0:None; 1:Minor; 2:Moderate; 3:Major; 4: Extreme)
SERVICES	Last service from the services list that a patient was admitted/transferred under. For the corresponding numeric values, in which '1' represents Cardiac Medical - for non-surgical cardiac related admissions, '2' represents Cardiac Surgery - for surgical cardiac admissions, … (The services , 2020).	Numeric	1–20
SERVICES_NUM	Number of services that a patient was admitted/transferred under	Numeric	1–20
DRG_COUNT	Number of DRG code for each patient	Numeric	1–76
COMORBIDITIES	If the subject has comorbidities or not	Boolean	Yes/No (1/0)
LOS	Length of stay (LOS) for the patient for the given ICU stay, which may include one or more ICU units. The length of stay is measured in fractional days.	Numeric	0–172
TRANSFER	The number of transfer to another hospital	Numeric	0–104
age_score	A score that represents the age of patients (Le Gall et al., 1993).	Numeric	0(0–39)/7(40–59)12(60–69)/15(70–74) 16(75–79)/18(>=80)0(0–27)/6(28–83)
bun_score	A score that represents blood urea nitrogen (Le Gall et al., 1993)	Numeric	10(>=84)
sysbp_score	A score that represents systolic blood pressure (Le Gall et al., 1993)	Numeric	0–13
comorbidity_score	A score that represents comorbidity (Le Gall et al., 1993)	Numeric	17(AIDS = 1) 10(HEM = 1) 9(METS = 1)/0(else)
acidbase_score	A score that represents acidbase (Le Gall et al., 1993)	Numeric	0–12
creatinine_score	A score that represents creatinine (Le Gall et al., 1993).	Numeric	0–10
temp_score	A score that represents inpatient temperature (Le Gall et al., 1993)	Numeric	0–20
uo_score	A score that represents urine output (Le Gall et al., 1993)	Numeric	0–11
gcs_score	A score that represents glasgow coma scale (Le Gall et al., 1993)	Numeric	0–26

different from patient to patient, one subject could have multiple medical records under each table content, others may only have one entry for several table contents. Therefore, it is essential to merge patient's multiple data records. For example, in the transfer table, there are multiple records for each subject. If a subject was admitted to a care unit, the EVENTTYPY will be marked with "admit"; if the subject transferred from one care unit to another, there will be one more record about the transfer information with the EVENTTYPY marked with "transfer"; if the subject discharged a care unit, the EVENTTYPE will be marked with "discharge". Based on database text querying, we synthesized several lines of records for one subject into one data entry, where one "admit" or "transfer" was counted as a one-time transfer. Although this table also provides information on previous care unit and current care unit names, this information may not be important at this stage. Similarly, how many times a patient has accepted medical services is calculated from the service table and became one feature as model input. The diagnose table provided detailed descriptions of the patient's illness diagnoses, which can be treated as the basis for defining whether the patient has comorbidity. For the issue we discussed, patients' information is much more crucial than the caregivers' information. Hence, the tables that provide information like the type of caregiver, such as the caregiver table, are excluded. Some tables have some information that cannot fit for the feature selection models. For instance, the service table documented the patient's medicine injection or taking information and the prescription table has various representations of the drug prescribed to the patient and the route prescribed for the drug.

To reflect a person's health status, using severity score is considered as a good way typically. Age and physiological indicators regard to heart rate, respiratory rate, blood pressure, temperature, etc. are significant variables that can be transformed into another presentation as scores. ICD9_code defined the type of comorbidities. Specific ITEMID in the Lab events or Chart events tables determined many physiological variables to get the severity score. Here we set scores for these features of ICU patients with two scoring standards, APS III and SAPS II. Acute Physiology Score (APS) was originally published alongside the Acute Physiology and Chronic Health Evaluation (APACHE) system by Knaus et al. at George Washington University (Knaus et al., 1981). The philosophy of APACHE system is that detailed information on the patient's acute severity of the illness can be assessed by the various physiologic data routinely measured on ICU patients (Wagner & Draper, 1984). Therefore, the system also contains an acute physiology score based on various physiologic measurements and a chronic health assessment. Each measurement is assigned by a score dependent on the quantile it rested within. Then the scores of different measurements are added together to get the ultimate score. The latest version of APS is the APS III, which is published in (Knaus et al., 1991) and used in our experiment. Simplified Acute Physiology Score (SAPS) only used 14 measured biologic and clinical variables, which are collected during the first 24 h after ICU admission (Le et al., 1984). Later on, in order to address the defects in SAPS, SAPS II was published (Le Gall et al., 1993) with two improvements. First, SAPS II

Table 2
Feature selection results of Weekdays GAM model and Weekends GAM model.

	Weekdays	Weekends
1	SERVICES	age_score
2	DRG_CODE	DRG_CODE
3	DRG_COUNT	comorbidity_score
4	age_score	SERVICES
5	SERVICES_NUM	temp_score
6	comorbidity_score	SERVICES_NUM
7	TRANSFER	COMORBIDITIES
8	bun_score	sysbp_score
9	DRG_MORTALITY	gcs_score
10	acidbase_score	acidbase_score
11	LOS	creatinine_score
12	COMORBIDITIES	uo_score
13	sysbp_score	DRG_COUNT
14	creatinine_score	DRG_SEVERITY

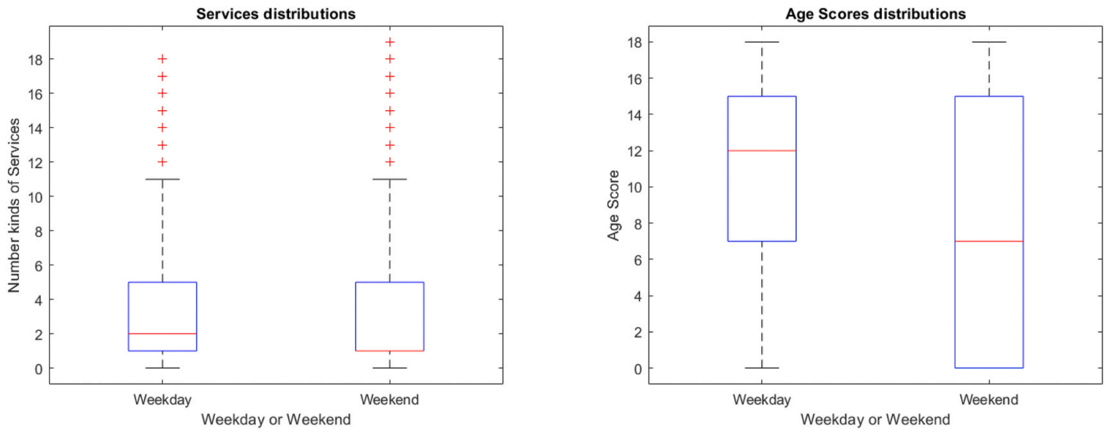

Fig. 2. Box plot analysis of two most significant features (Services and age_score) identified by the Weekday GAM and Weekend GAM models separately. Red line indicates the median; blue box indicates the interquartile range, where upper boundary is 75th percentile and bottom boundary is 25th percentile; black line indicates the maximum and minimum; red crosses indicate outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

utilized the univariate analysis to select features that were correlated with hospital mortality. Second, SAPS II also provided a method to convert the score to a probability of hospital mortality.

Particularly, we try to avoid omitting any information and any subjects provided by this medical dataset. Some tables that do not contain subject ID are not considered in this study, because the final input file for the models should be generated by querying subject ID in each separate tables and recombination. Columns only have the identification number or specific date and time of a kind of event are not considered, because a single ID number or an event time-point is lack of analyzing values to the outcomes. On the other hand, 559 subjects that lack too many records of features are not considered. After the data screening, 11 of 26 tables were identified with a total of 46,476 patients that can be used for this study. Since the input and output data formats have to be numeric for the feature selection models, we have to transform some text information into numerical expressions, such as using '1' express Male, '2' express Female. Then 41 features were generated finally 1. Moreover, the ground-truths were binary, flagged as alive and death for two separate models.

5. Experiment and evaluation

5.1. Finding predominant factors that leading to death for weekday or weekend admission to ICU

In terms of the weekday admission for a patient and the weekend admission for a patient are two independent events, we construct two GAM models: the Weekday GAM model and the Weekend GAM model. If we could find the critical features for each model first, i.e. the predominant factors associated with death for weekday admission or weekend admission, we could analyze the mortality difference by studying the two models' feature selection results. We apply Algorithm 1 for feature elimination and refit the models at each iteration. In this way, the feature rankings are keeping updated. At each iteration, we found that the overall output ranking will change a little or not. This can be explained as the selection bias as we mentioned earlier. Therefore, this is the reason why Algorithm 1 is

Table 3
Analysis for some selected features. Mean and standard deviation are listed in the table. More times of transfer and number of services occurred for Weekday admission. The greater standard deviations of Weekend admission indicate weekend data fluctuate violently.

	Weekdays	Weekends
Patients Admissions	36240	10236
TRANSFER	4.42 (1.05)	4.16 (3.74)
SERVICES_NUM	1.58 (0.37)	1.54 (1.32)
DRG_MORTALITY	1.44 (0.39)	1.51 (1.58)
temp_score	1.03(0.48)	0.91 (2.33)

Table 4
Classification results for weekdays and weekends mortality with bagged tree classifier.

Feature eliminations	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Overall (Weekdays)	79.70	78.29	78.93	81.85
Overall (Weekends)	79.15	78.74	78.94	80.64
GAM (Weekdays)	77.65	76.61	77.13	79.90
GAM (Weekends)	76.84	76.87	76.86	78.56
PCA (Weekdays)	76.14	74.46	75.29	78.47
PCA (Weekends)	76.17	75.57	75.87	77.89
SAPS II (Weekdays)	70.39	65.66	67.94	73.14
SAPS II (Weekends)	71.31	70.25	70.78	73.46
APS III (Weekdays)	67.80	63.40	65.52	71.31
APS III (Weekends)	69.33	63.94	66.53	70.78

implemented as well. After some iterations, the final feature rank for each model is identified, and 14 most significant features are listed as well. The best list of features from the weekday admission model are SERVICES, DRG_CODE, DRG_COUNT, age_score, SERVICES_NUM, etc. The best list of features from the weekend admission model are age_score, DRG_CODE, comorbidity_score, SERVICES, etc. Table 1 and Table 2 summarized the top 14 feature rankings of weekday admission analysis and weekend admission analysis.

By comparing the feature ranking from the Weekday GAM model with that from the Weekend GAM model, we found that 10 features are included in both models top rankings, which means these 10 features are very important to evaluate the patient's possibility of death or alive that independent from the day of admission. However, there are 4 different features in the mortality predictions of the two GAM models. Particularly, the first most significant features of the two models are inconsistent. Fig. 2 exhibits the box plot analysis of the two most crucial features identified by the Weekday GAM and Weekend GAM model separately. From this figure, weekend ICU admissions have less number of types of services comparing to weekday ICU admissions. This might be poorer instant services given by caregivers for weekend admissions. The age_score difference indicates that weekend ICU admissions have a wider range of age groups of acute patients, which may indirectly reflect the possible situation that more acute illness may impel the patients to have to visit ICU on weekends. Then we did statistical analysis on selected four other significant features in Table 3. Weekday admissions present more times of transfer and a greater number of services occurred. Moreover, weekend variable standard deviations are much greater than weekdays that indicates weekend data fluctuate violently. Based on the feature ranking, the problems of weekend admitted patients' uneven temperatures and higher DRG mortality are worth of medical providers' attention.

5.2. *Evaluate the effectiveness of the feature selection method*

In this section, we are going to evaluate if the GAM models have successfully found the critical features for each specific problem: causes of weekday admission death; causes of weekend admission death. The number of survivors is greater than the number of deaths for both weekday admission and weekend admission, meaning. This imbalance makes the training classifier biased inevitably. Hence, to balance the dataset, the amount of death data that is less than the amount of survival data is filled with available death data by the method of overall data samples upsampling and residuals random upsampling. Then the bagged tree machine learning method is selected. Algorithm 1 comprised a 4-fold cross-validation strategy with contrast experiments that are reported to evaluate the performance of the proposed feature selection results. Table 4 shows the classification results.

For the weekday admission prediction, the accuracy is 81.85% when using a total of 41 features. For the weekend admission prediction, the accuracy is 80.64% when a total of 41 features are applied. When the number of features that are used for training is reduced to 14, the predicting.

accuracy is still 79.90% for the weekday model and 78.56% for the weekend. In the contrast experiments as shown in Table 4, three other feature elimination methods are listed: PCA (principal components analysis), SAPSII, and APSIII which are introduced in the "Data Preparation" stage.

Four metrics that present the fraction of the patients who are correctly predicted as passed-away over the whole number of passed-away patients are used to evaluate the performances of the feature selections. Except for the prediction accuracy, the other three

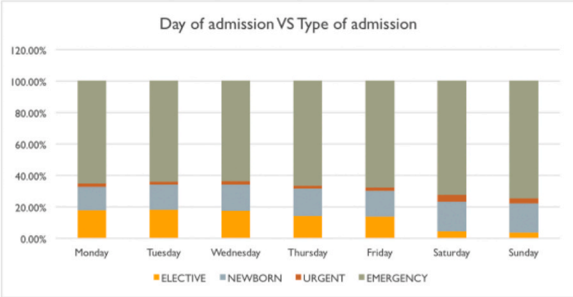

Fig. 3. Emergency/urgent indicate unplanned medical care. Elective indicates a previously planned hospital admission. Newborn indicates that the HADM_ID pertains to the patient's birth. More percentage of unplanned medical cares (URGENT and EMERGENCY) enter ICU during weekend than that of weekday.

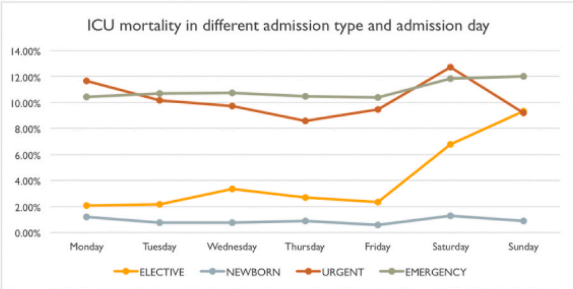

Fig. 4. This statistical results indicate that urgent and emergency admissions are more likely to die. Saturday admission to ICU usually has higher probability of death than other days.

metrics are precision, recall, and F1-score as shown in equation (4). These metrics calculate the quality of classification for weekdays admissions and weekend admissions.

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1score &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \tag{4}$$

where "TP" is the patients who are truly diagnosed as passed away, "FP" is the ones who are incorrectly labeled as passed away, "FN" is the ones who are incorrectly labeled as living patients. The GAM method outperforms all the other feature selections which are easily interpretable and provide some clinical insights into the classification process. Meanwhile, the values for F1-score among the transparent classifiers demonstrate a big gap between the decision tree and the others. Whereas the F1-score of SAPSII, and APSIII is around "70%" and the PCA's is "76%", the GAM has a higher F1-score of "77%". Furthermore, the other three methods are constructed through multi-features fusion which loses interpretability. The contrast experiments have proved that our proposed GAM feature selection method is effective in discovering the predominant factors associated with the death of patients admitted on weekday or weekend.

5.3. Difference between patients of weekday admission and weekend admission exists in same features

In this experiment, the bagged tree is trained by the weekday data and tested with the weekend data; vice versa, it is trained by the weekend data and tested with the weekday data. Compared with the last experiments' prediction accuracies that 81.85% of 41 features weekday admission prediction and 80.64% of 41 features weekend admission prediction, the prediction accuracies are getting worse at this time: weekday admission prediction accuracy is only 58.08% and weekend admission prediction accuracy is just 63.98% with overall 41 features. It indicates that there is an obvious difference between the patient data of weekday admission and that of weekend admission. This observation demonstrates that the difference for common factors under two different models contributes to the "weekend effect". Typically, the admission type is selected as an example for statistical analysis to prove the discrepancy. The entire database is divided into four categories based on types of admission, which are Newborn, Elective, Urgent, and Emergency. Emergency/urgent indicate unplanned medical care and are often collapsed into a single category in studies. Elective indicates a previously planned hospital admission. Newborn indicates that the subject admission pertains to the patient's birth. Fig. 3 showed a higher percent of unplanned medical care is accepted on weekends. As we expected, the statistical results in Fig. 4 just prove that urgent and emergency admissions have higher mortality than the other two types. Moreover, it is interesting to see that elective admissions have almost consistent mortality of the five weekdays, but this mortality has increased apparently on Saturday and Sunday. The experiment has demonstrated that the same features' variables are different from weekday admission to weekend admission.

6. Conclusion

The statistical result according to the MIMIC III database showed that there is an obvious discrepancy in mortality for patients admitted to the intensive care unit on weekends and those admitted on weekdays. We have built two GAM models to find the predominant factors associated with death for patients admitted on weekday or weekend, and significant features were identified by the methods. An evaluation experiment has proved the reliability of the GAM feature selection method, which overwhelms the other feature elimination methods and has high interpretability. Furthermore, although more common significant features are identified in both models, we also found specific variables of predominant factors associated with the death of patients admitted on weekdays are not identical to that on weekends. We are looking forward that our findings could provide more evidence on supporting medical service providers to tackle medical problems in the future.

CRediT authorship contribution statement

Yi Cai: Conceptualization, Methodology, Software, Writing - original draft. **Jianian Zheng:** Methodology, Software, Writing - original draft. **Xiaoliang Zhang:** Methodology, Software, Investigation. **Haotian Jiang:** Writing - review & editing. **Ming-Chun Huang:** Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank the MIT Lab for Computational Physiology, Massachusetts Institute of Technology for the experimental database, and IEEE Conference on Biomedical and Health Informatics (BHI) 2018 and the IEEE Conference on Body Sensor Networks (BSN) for providing a platform to discuss this problem.

References

Barba, R., Losa, J., Velasco, M., Guijarro, C., de Casasola, G. G., & Zapatero, A. (2006). Mortality among adult patients admitted to the hospital on weekends. *European Journal of Internal Medicine, 17*(5), 322–324.

Barnett, M. J., Kaboli, P. J., Sirio, C. A., & Rosenthal, G. E. (2002). Day of the week of intensive care admission and patient outcomes: A multisite regional evaluation. *Medical Care, 40*(6), 530–539.

Bell, C. M., & Redelmeier, D. A. (2001). Mortality among patients admitted to hospitals on weekends as compared with weekdays. *New England Journal of Medicine, 345*(9), 663–668.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Calvert, J., Mao, Q., Hoffman, J. L., Jay, M., Desautels, T., Mohamadlou, H., Chettipally, U., & Das, R. (2016). Using electronic health record collected clinical variables to predict medical intensive care unit mortality. *Annals of medicine and surgery, 11*, 52–57.

Celi, L. A., Zimolzak, A. J., & Stone, D. J. (2014). Dynamic clinical data mining: Search engine-based decision support. *JMIR medical informatics, 2*(1).

Clark, M.. Generalized additive models. http://m-clark.github.io/docs/GAM.html#generalized_linear_model.

Ensminger, S. A., Morales, I. J., Peters, S. G., Keegan, M. T., Finkielman, J. D., Lymp, J. F., & Afessa, B. (2004). The hospital mortality of patients admitted to the icu on weekends. *Chest, 126*(4), 1292–1298.

Fontaine, P. (2004). Using severity adjustment classification for hospital internal and external benchmarking. *Using severity adjustment classification for hospital Internal and External Benchmarking/AHIMA*. American Health Information Management Association.

Freemantle, N., Ray, D., McNulty, D., Rosser, D., Bennett, S., Keogh, B. E., & Pagano, D. (2015). Increased mortality associated with weekend hospital admission: A case for expanded seven day services? *British Medical Journal, 351*.

Freemantle, N., Richardson, M., Wood, J., Ray, D., Khosla, S., Shahian, D., Roche, W., Stephens, I., Keogh, B., & Pagano, D. (2012). Weekend hospitalization and additional risk of death: An analysis of inpatient data. *Journal of the Royal Society of Medicine, 105*(2), 74–84.

James, M. T., Wald, R., Bell, C. M., Tonelli, M., Hemmelgarn, B. R., Waikar, S. S., & Chertow, G. M. (2010). Weekend hospital admission, acute kidney injury, and mortality. *Journal of the American Society of Nephrology, 21*(5), 845–851.

Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific data, 3*, 160035.

Knaus, W. A., Wagner, D. P., Draper, E. A., Zimmerman, J. E., Bergner, M., Bastos, P. G., Sirio, C. A., Murphy, D. J., Lotring, T., Damiano, A., et al. (1991). The Apache iii prognostic system: Risk prediction of hospital mortality for critically iii hospitalized adults. *Chest, 100*(6), 1619–1636.

Knaus, W. A., Zimmerman, J. E., Wagner, D. P., Draper, E. A., & Lawrence, D. E. (1981). Apache-acute physiology and chronic health evaluation: A physiologically based classification system. *Critical Care Medicine, 9*(8), 591–597.

Le Gall, J.-R., Lemeshow, S., & Saulnier, F. (1993). A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. *Jama, 270*(24), 2957–2963.

Le, J. G., Loirat, P., Alperovitch, A., Glaser, P., Granthil, C., Mathieu, D., Mercier, P., Thomas, R., & Villers, D. (1984). A simplified acute physiology score for icu patients. *Critical Care Medicine, 12*(11), 975–977.

Lior, R., et al. (2014). *Data mining with decision trees: Theory and applications, ume 81*. World scientific.

Lou, Y., Caruana, R., & Gehrke, J. (2012). Intelligible models for classification and regression. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM.
Meacock, R., Anselmi, L., Kristensen, S. R., Doran, T., & Sutton, M. (2017). Higher mortality rates amongst emergency patients admitted to hospital at weekends reflect a lower probability of admission. *Journal of Health Services Research & Policy, 22*(1), 12–19.
Miles, J. (2014). *R squared, adjusted r squared*. Wiley StatsRef: Statistics Reference Online.
S. Ranji, R. J. Stern, R. Koppel, and U. Sarkar. The weekend effect.
Sadeghi, R., Banerjee, T., & Romine, W. (2018). Early hospital mortality prediction using vital signals. *Smart Health, 9*, 265–274.
The drgcodes table. (2020). https://mimic.physionet.org/mimictables/drgcodes/.
The services table. (2020). https://mimic.physionet.org/mimictables/services/.
Wagner, D. P., & Draper, E. A. (1984). Acute physiology and chronic health evaluation (Apache ii) and medicare reimbursement. *Health Care Financing Review*, (Suppl), 91, 1984.
Walker, A. S., Mason, A., Quan, T. P., Fawcett, N. J., Watkinson, P., Llewelyn, M., Stoesser, N., Finney, J., Davies, J., Wyllie, D. H., et al. (2017). Mortality risks associated with emergency admissions during weekends and public holidays: An analysis of electronic health records. *The Lancet, 390*, 62–72, 10089.
Wikipedia. (). *Diagnosis-related group — wikipedia, the free encyclopedia*. http://en.wikipedia.org/w/index.phptitle=Diagnosis/related/groupoldid=977693664 [Online; accessed 20-October-2020].
Wikipedia. (2020a). F-test — wikipedia, the free encyclopedia [Online; accessed http://en.wikipedia.org/w/index.php?title=F-test&oldid=961881016. (Accessed 17 September 2020).