Received February 12, 2022, accepted March 8, 2022, date of publication March 21, 2022, date of current version March 29, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3160742*

# Analysis on Benefits and Costs of Machine Learning-Based Early Hospitalization Prediction

**EUNBI KIM[1], KAP SU HAN[2], TAESU CHEONG[1], SUNG WOO LEE[2], JOONYUP EUN[3], AND SU JIN KIM[2]**

[1]School of Industrial and Management Engineering, Korea University, Seoul 02841, South Korea
[2]Department of Emergency Medicine, College of Medicine, Korea University, Seoul 02841, South Korea
[3]Graduate School of Management of Technology, Korea University, Seoul 02841, South Korea

Corresponding authors: Joonyup Eun (jeun@korea.ac.kr) and Su Jin Kim (icarusksj@korea.ac.kr)

This work was supported in part by Korea University Anam Hospital, Seoul, Republic of Korea, under Grant No. K1912671; and in part by the National Research Foundation of Korea (NRF) grant funded by Korea Government, Ministry of Science and ICT, under Grant No. 2020R1F1A1075832.

**ABSTRACT** Overcrowding in emergency departments (EDs) has long been a problem worldwide and has serious consequences for patient satisfaction and safety. Typically, overcrowding is caused by delays in the boarding time of ED patients waiting for inpatient beds. If the hospitalization of patients is predicted early enough in EDs, inpatient beds can be prepared in advance and the boarding time can be reduced. We design machine learning-based hospitalization predictive models using data on 27,747 patients and compare the experimental results. Five predictive models are designed: 1) logistic regression, 2) XGBoost, 3) NGBoost, 4) support vector machine, and 5) decision tree models. Based on the predictive results, we estimate the quantitative effects of hospitalization predictions on EDs and wards. Using the data from the ED of a general hospital in South Korea, our experiments show that the ED length of stay of a patient can be reduced by 12.3 minutes on average and the ED can reduce the total length of stay by 340,147 minutes for a year.

**INDEX TERMS** Emergency department, machine learning, hospitalization prediction, estimation of quantitative effects.

## I. INTRODUCTION

Emergency department (ED) overcrowding is a severe problem in the health sector worldwide [1]. It occurs when a discrepancy is observed between the medical demands of ED patients and resource supplies that are required for proper patient flow and treatment [2]. ED overcrowding reduces the quality of treatment for patients and increases their length of stay (LOS) and mortality rate [2], [3]. The resources related to handling ED overcrowding include treatment-related personnel, testing laboratories, and inpatient beds. The lack of available beds to accommodate patients hospitalized in EDs is the most critical factor for ED overcrowding [4].

Boarding time is defined as the time between making the clinical decision to hospitalize ED patients and their departure from the ED. This time is often prolonged because the demand for inpatient beds outweighs the availability of beds [5], [6]. During the fourth quarter of 2018, the average LOS for hospitalized patients in the ED at a general hospital in South Korea was more than 7 h (Fig. 1).

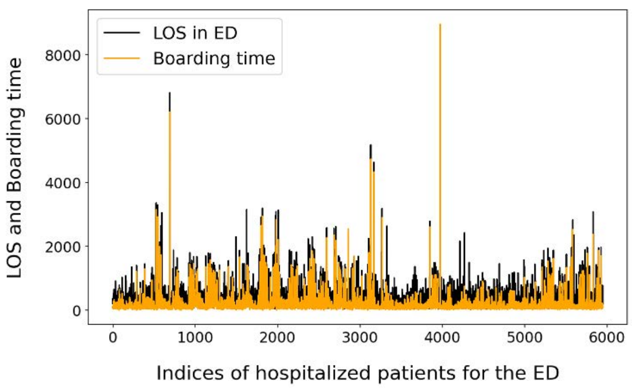



**FIGURE 1.** Ratio of boarding time to LOS in the ED (fourth quarter of 2018, Korea University Anam Hospital).

The associate editor coordinating the review of this manuscript and approving it for publication was Dominik Strzalka.

The ratio of the boarding time to the LOS in the ED was 47% on average. Since boarding time is a hindrance to receive medical treatment in the ED, it is beneficial to reduce the boarding time as much as possible. However, the competition between ED patients and outpatients for a finite number of inpatient beds increases the boarding time.

Predicting the hospitalization of ED patients is one of the measures taken to reduce the boarding time and facilitate inpatient bed management, staff planning, and specialized workflows within the ED [7]. This study hypothesizes that hospitalization predictions can initiate the preparation of inpatient beds in advance and ultimately help reduce the LOS of ED patients. Therefore, we aim to identify a model that accurately predicts ED patients who are hospitalized to inpatient beds at an early stage of ED stay. We also estimate the quantitative effects of hospitalization predictions on EDs and wards and the extent to which they contribute to reducing the LOS in the ED.

We performed a predictive analysis for a single general hospital. ED patients' flows are similar across general hospitals. ED patients typically go through the following steps: ED entrance, triage and initial exam, treatment, disposition, and hospitalization or discharge from the ED [8], [9]. In addition, most EDs obtain similar clinical information from the initial exams for their patients [10], [11]. This study uses data recorded from the ED patients' flow and initial exams that are similarly implemented at general hospitals. For these reasons, there is little complication in applying the machine learning methods and quantitative effect analysis to their hospitals.

The remainder of this paper is organized as follows. Section II reviews various algorithms and previous literature on hospitalization predictions for ED patients. The data, algorithms, and experimental settings for this study are described in Section III. Section IV explains the experimental results. Section V discusses the interpretation of the quantitative effects of hospitalization predictions. Finally, the conclusions of this study are presented in Section VI.

## II. RELATED LITERATURE

This section introduces various predictive studies for ED patients. Some researchers have generally used logistic regression (LR) and ensemble-based classification algorithms in predictive studies. Kim *et al.* [12] predicted the hospitalization of patients visiting the ED and showed which characteristics of the patients influenced their likelihood of hospitalization. Their predictive model was mainly analyzed based on *accuracy* and the area under the ROC curve (*AUC*). They found that the older the patient and the more urgent their condition, the more likely they were to be hospitalized. Lucke *et al.* [13] divided 21,287 ED patients into two groups (>70 years and <70 years old) and predicted hospitalization using LR. They evaluated patients' hospitalization predictions using indices like *AUC and* positive prediction values. Their study demonstrated that predictive models could help identify patients who were more likely to be hospitalized using readily available information, such as their vital signs. Graham *et al.* [11] used three algorithms, namely LR, the gradient boosting model (GBM), and decision tree (DT), to predict the hospitalization for ED patients and analyzed 107,545 patient data. They suggested that when choosing a predictive model, simplicity and interpretation efficiency took precedence over the model's performance. Some studies have considered using neural networks in addition to regression and ensemble-based classifications for hospitalization predictions. Araz *et al.* [14] performed hospitalization predictions based on LR, DT, support vector machine (SVM), extreme gradient boosting (XGBoost), random forest (RF), and artificial neural network (ANN) models using data from 118,005 patients. Among the predictive models, XGBoost showed the highest *AUC*. Hong *et al.* [8] analyzed LR, XGBoost, and deep neural networks (DNNs). Based on 560,486 patient visits, they analyzed three groups of data: patient severity classification data, clinical data from previous visits, and all the available data from previous and current visits. In this analysis, XGBoost and DNN displayed good *AUC* values when predicting ED patient hospitalizations. Golmohammadi [15] presented hospitalization predictions using LR, ANN, and a statistical method that patterned the similarity of patient characteristics to predict hospitalization. He showed that the overall *accuracy* of the three models was greater than 80%.

ED data include information on the main symptoms of the patients who visit the ED. In general, these symptoms are recorded in free text and are not standardized. Even the same symptom is expressed in various ways. Consequently, the number of main symptoms in the data is quite large, reaching hundreds or thousands [16]. Since including several symptoms in a predictive analysis could reduce the prediction accuracy, if the target of the analysis was narrowed to a specific patient (e.g., a diabetic patient), the number of main symptoms that needed to be preprocessed would decrease. Therefore, the complexity of preprocessing can be reduced, and the *accuracy* of the predictions can increase. Accordingly, some researchers have performed predictive studies restricting the target of the analysis, such as studies on patients with a specific disease or patients of a particular age group. Dinh *et al.* [17] limited the targets of analysis to adult patients aged 16 years or older and included 860,832 patient data in the analysis. LR was used to predict hospitalizations to improve the patient flow and aid clinical decision-making in the ED. LR was interpreted based on *AUC*. Their study showed that accurate hospitalization predictions for ED patients could be made using initially available patient information, such as age, mode of arrival, and time of arrival. Fenn *et al.* [18] constructed a predictive model using LightGBM to divide the likelihood of the hospitalization into four stages: low, medium, high, and very high. A total of 468,167 patient data points were used. Medical personnel could respond flexibly to patients' follow-up processes by dividing them into several categories according to their likelihood of hospitalization. This constructed predictive model was measured based on *AUC*. Goto *et al.* [19] studied hospitalization predictions for children who visited the ED. Using data from 52,037 children, they used lasso regression, RF, XGBoost, and DNN to predict two clinical outcomes: critical care or hospitalization. These models were evaluated for their *sensitivity* and *specificity*; DNN was the best predictor of hospitalization in children. Horng *et al.* [20] used data from 230,936 ED patients to predict their diseases. In addition to curated data, such as patients' vital signs, they used free-text information on important symptoms to identify patients with sepsis. They employed an SVM to predict sepsis and assessed it based on *AUC*. The SVM achieved high performance using data routinely available during triage (e.g., reasons for visits and vital signs). Ram *et al.* [21] conducted a study to predict the number of daily ED visits of asthmatics using Twitter and Google data collected from various regions. They showed that a predictive model using data over a short period of three months could predict the ED visits of asthmatics in near real-time. The DT and ANN models that were used predicted the number of ED visits for asthmatics in daily low-, medium-, and high-volume categories. The models were evaluated in terms of their *AUC* and precision values; the predictive *accuracy* of the medium-volume category obtained using the hybrid ANN and DT models was the highest. Barack-Corren *et al.* [7] studied the prediction of hospitalization for pediatric patients who visited the ED. A total of 59,033 patient data points were used, and the predictions were tested using the data that were available within 10, 30, and 60 min after the patients arrived at the ED. The predictions were made using a hybrid model that combined Naive Bayes and LR. They estimated the potential effects of hospitalization predictions for ED patients. From the perspective of the ED patients' flow, they derived the effects of the *time saved in the ED* and *the time costs in the inpatient ward* (i.e., the total time during which empty beds were held for ED patients in the inpatient ward).

Natural gradient boost (NGBoost) is a boosting-based machine learning algorithm published in 2019 by Duan *et al.* [22]; this algorithm was designed to estimate the uncertainty in regression prediction, such as in probabilistic precipitation prediction. They showed that NGBoost offers competitive performance in negative log-likelihood, particularly when working on small datasets. In healthcare research, NGBoost has been used for predictive studies based on machine learning, such as brain tumor predictions [23] and treatment frequency predictions for macular degeneration [24]. However, because NGBoost has not been used for ED hospitalization predictions, this study intends to investigate its use in hospitalization predictions for ED patients.

As machine learning research has become increasingly dynamic, hospitalization predictions for ED patients are being steadily studied.

Table 1 summarizes machine learning-based predictive studies for ED patients. Except for the studies of Jaccinta (2016) and Davood (2016), most studies have used a large amount of data, ranging from data on 50,000 to that on 500,000 cases. However, in practice, medical data collection faces many challenges, owing to patient privacy and organizational issues. In addition, if a large amount of data is available, a large amount of time and computing resources are required to train the models on the data. For these reasons, many research cases may confine the use of large datasets [25].

**TABLE 1. Predictive studies for ED patients.**

| Authors | Samples (≤ 50,000) | Predictive algorithms | Analysis targets (All ED patients) | Verification of the predicted effectiveness |
|---|---|---|---|---|
| KIM et al. [12] | | LR | ○ | |
| Lucke et al. [13] | ○ | XGBoost | | |
| Graham et al. [11] | | LR, GBM, DT | ○ | |
| Araz et al. [14] | | LR, DT, SVM, XGBoost, RF, ANN | ○ | |
| Hong et al. [10] | | LR, XGBoost, DNN | ○ | |
| Golmohammadi [15] | ○ | LR, ANN | ○ | |
| Dinh et al. [17] | | LR | | |
| Fenn at al. [18] | | LightGBM | | |
| Goto at al. [19] | | RF, XGBoost, DNN | | |
| Horng et al. [20] | | SVM | ○ | |
| Ram et al. [21] | | DT, ANN | | |
| Barack -Corren et al. [7] | | Naive Bayes, LR | | ○ |
| Our study | ○ | LR, XGBoost, SVM, DT, NGBoost | ○ | ○ |

However, if the quantity of collected data is smaller, hospitalization predictions may be less effective.

This study makes four major contributions. First, this study shows that high-*accuracy* hospitalization predictions are possible using data on less than 50,000 patients. In our study, we employ 27,747 patient data points, which are significantly less than those of previous studies, to predict hospitalization for ED patients. This study also verifies predictive performance based on the amount of data. Second, this study analyzes hospitalization predictions for all ED patients without confining the group under study to patient characteristics such as age and disease. The long boarding time of a patient between their stay in an ED and that in a ward contributed to ED overcrowding, regardless of each patient's specific disease and age. Third, this study uses NGBoost, which has not been used for ED patient hospitalization predictions. We use a smaller set of data than that used in previous studies. Therefore, we need an algorithm suitable for deriving prediction performance from a small set of data. Because NGBoost has the advantages which derive a good prediction performance from a small set of data, according to Duan *et al.* [22], we use it. Finally, this study shows the quantitative effects of hospitalization predictions on EDs and wards, which most studies have not. This study refers to the method used by

Barack-Corren *et al.* [7] to estimate the effects of hospitalization predictions. We reconstruct the method according to the flow of ED patients and the administrative system and demonstrate how much hospitalization predictions could reduce ED overcrowding.

## III. MATERIALS AND METHODS

The data collected from the ED include all the electronic records of the patient, from their date of visit to the reason for their hospitalization or discharge [26]. This section describes the information on the variables contained in the data and the design of predictive models for hospitalization predictions. Furthermore, this section explains a method for estimating the *time saved in the ED* and *the time costs in inpatient ward* when hospitalization predictions are adopted in practice.

### A. DATA COLLECTION AND DESCRIPTION

This study is based on data recorded in the electronic system of Korea University Anam Hospital (KUAH), a general hospital in South Korea. The collected data include records on 27,747 patients for seven months from October 2018 to April 2019. An average of 131 patients visit the hospital's ED per day, and approximately 26 ED patients (20%) are admitted to the general wards or intensive care units. Excepting personal patient information, such as names and medical record numbers, we use all the collected data on the ED patient records. This study uses 22 variables collected within 20 min after a patient's entrance to the ED. Within this timeline, the following information is available:

#### 1) DEMOGRAPHICS

It refers to basic information such as the patient's age, gender, and time of arrival at the ED. Gender variables are recorded in eight categories for each gender of the patient, based on their year of birth (the 1900s and 2000s) and nationality (Korean and foreigner).

#### 2) DISEASE STATUS

This refers to information about whether a patient has a disease or not. The disease's status is classified into three categories: presence, absence, and others.

#### 3) VISIT ROUTE

This variable refers to the place from where a patient arrives at the ED. The patient may have arrived directly or have been transferred to the ED from another hospital.

#### 4) CATEGORY OF INDIVIDUALS IN TRAFFIC ACCIDENTS

This variable describes a patient's role at the time of an accident (e.g., whether the patient was a pedestrian or a driver).

#### 5) INDIVIDUALS IN TRAFFIC ACCIDENT

In the case of a car accident, these variables describe the situation at the time of the accident (e.g., whether the patient is wearing a helmet or knee protector).

#### 6) ARRIVAL MODE

This variable refers to the transportation mode used by the patient to reach the ED. It is classified into walking, hospital ambulance, public transportation, or others.

#### 7) TRIAGE

Triage evaluates and prioritizes the severity of a patient's injury or illness within a short period of time after the patient arrives at the ED. The closer the triage is to level 1, the more severe the patient's status is.

#### 8) REACTION STATUS

This describes how a patient reacts when arriving at the ED. It is divided into five categories, depending on the patient's reaction.

#### 9) CHIEF COMPLAINTS

The chief complaints are the patient's symptoms when the patient arrives at the ED, such as abdominal pain and fever. These chief complaints are collected from texts and include 379 symptoms. Most complaints that occurred in the bottom 5% frequency are observed only once. These are all categorized as "others."

#### 10) VITAL SIGNS

Vital signs are important indicators of the state of a body's life support functions. The ED staff quickly checks for signs of patient vitality before seeing the patient. Blood pressure, body temperature, pulse rate, oxygen saturation, and breathing are five vital signs observed in the ED.

#### 11) CAUSE OF INJURY

The cause of injury is any physical or chemical source that leads to injury (e.g., falls, slips, burns, and drug addiction).

#### 12) INTENTIONALITY OF INJURY

The intentionality of injury refers to whether a patient's injury occurs intentionally or not.

#### 13) HOSPITALIZATION

Hospitalization is the dependent variable in this study. This explains whether the patient is hospitalized or not.

Categorical variables are converted to binary variables using one-hot encoding [27]. After preprocessing, 220 variables are used for the predictive models, and there are no missing values for the 27,747 ED patients.

Table 2 summarizes the basic descriptive statistics of each variable for the 27,747 patients. The $p$-value of each variable represents the statistical significance of hospitalization. As summarized in Table 2, the ages of the patients in the ED are evenly distributed. Patients in their 90s or older have the highest likelihood of hospitalization (48.2%). A total of 73.1% of patients with ED have diseases, of which 26% are hospitalized. This accounts for 89.9% of the 5,949 ED inpatients and is consistent with the results of the study that patients with chronic diseases, such as diabetes and sepsis, frequently visited the ED [28]. A total of 0.1% of ED patients use hospital ambulances to visit the ED, and 80.8% go to hospitalization wards. On the contrary, 51.2% of ED patients are assigned to triage level 3, accounting for the largest proportion. However, the hospitalization ratio of patients is in the decreasing order of the triage level (i.e., the hospitalization ratio of level 1 patients is the highest and that of level 5 patients is the lowest).

### B. MACHINE LEARNING ALGORITHMS

Most studies related to hospitalization prediction have used machine learning algorithms. In this study, we also use machine learning algorithms to classify ED patient hospitalization; 1 for hospitalization and 0 for discharge from the ED. Since LR, SVM, and DT have provided good hospitalization predictions [11], [14], we use them in our study as well. Additionally, since XGBoost is known to be superior to other algorithms in terms of generalization performance and *accuracy* in several fields [29], we predict hospitalization using XGBoost. We include NGBoost, which is a recent algorithm that has not been used for ED hospitalization predictions in other experiments.

#### 1) LOGISTIC REGRESSION (LR)

LR is an efficient and straightforward method for binary or multiple classification problems. It uses the logit or natural log of the odds so that the probability of the data belonging to a particular class is not excluded from the [0, 1] range. LR is a supervised learning algorithm that categorizes classes according to probability and provides accurate predictions [30].

#### 2) EXTREME GRADIENT BOOSTING (XGBOOST)

XGBoost is a highly scalable algorithm developed to improve performance and computational speed. Boosting is an ensemble technique that adds new models to accommodate for errors made by existing models. Gradient boosting is used to create new predictive models using the residuals of fitted models and minimize losses. XGBoost can be used for both regression and classification [31].

#### 3) SUPPORT VECTOR MACHINE (SVM)

SVM is a linear learning method and classification method in supervised learning that finds the optimal hyperplane that separates two classes. It maximizes the distance between the two closest classes to achieve a high classification performance [32]. The data points for the two classes closest to the determined decision boundary are called the support vectors. The distance between the support vector and decision boundary is called the margin, and the decision boundary that maximizes the margin is optimal [33].

#### 4) DECISION TREE (DT)

DT is a nonparametric supervised learning method that is used for classification and regression. It implements a simple

**TABLE 2. Predictive variables and outcomes for 27,747 ED patients.**

| | Variables | Categories | Cases | Hospitalization | P |
|---|---|---|---|---|---|
| 1) | Age | | | | < .001 |
| | | < 30 | 8,690 (31.3%) | 854 (9.8%) | |
| | | 30–59 | 9,718 (35%) | 1,727 (17.8%) | |
| | | 60–89 | 9,028 (32.5%) | 3,218 (35.6%) | |
| | | ≥ 90 | 311 (1.1%) | 150 (48.2%) | |
| 2) | Gender | | | | < .001 |
| | | Male-Local Residents | 13,416 (48.4%) | 3,231 (24.1%) | |
| | | Male- Foreign Residents | 299 (1.1%) | 40 (13.4%) | |
| | | Female- Local Residents | 13,629 (49.1%) | 2,641 (19.4%) | |
| | | Female- Foreign Residents | 403 (1.5%) | 37 (9.2%) | |
| 3) | Disease status | | | | < .001 |
| | | Diseased | 20,289 (73.1%) | 5,349 (26.4%) | |
| | | Not diseased | 7,458 (26.9%) | 600 (8.0%) | |
| 4) | Visit route | | | | < .001 |
| | | Direct visit | 23,017 (83%) | 3,995 (17.4%) | |
| | | Transfer | 3,344 (12.1%) | 1,675 (50.1%) | |
| | | Request from the OPD | 380 (1.4%) | 278 (73.2%) | |
| | | Others | 1,006 (3.6%) | 1 (0.1%) | |
| 5) | Individuals in traffic accident | | | | 0.015 |
| | | Driver | 533 (1.9%) | 54 (10.1%) | |
| | | Passenger | 198 (0.7%) | 17 (8.6%) | |
| | | Pediatrician | 305 (1.1%) | 38 (12.5%) | |
| | | Unknown | 1 (0.0%) | 1 (100.0%) | |
| | | Others | 26,710 (96.3%) | 5,839 (21.9%) | |
| 6) | Seatbelt | | | | < .001 |
| | | Yes | 235 (0.8%) | 13 (5.5%) | |
| | | No | 27,512 (99.2%) | 5,936 (21.6%) | |
| 7) | Child seat | | | | 0.641 |
| | | Yes | 5 (0.02%) | 1 (20.0%) | |
| | | No | 27,742 (99.98%) | 5,948 (21.6%) | |
| 8) | Airbag | | | | 0.999 |
| | | Yes | 7 (0.03%) | 2 (28.6%) | |
| | | No | 27,740 (99.97%) | 5,947 (21.4%) | |
| 9) | Helmet | | | | < .001 |
| | | Yes | 283 (1%) | 32 (11.3%) | |
| | | No | 27,464 (99%) | 5,917 (21.5%) | |
| 10) | Joint protector | | | | 0.641 |
| | | Yes | 5 (0.02%) | 1 (20.0%) | |
| | | No | 27,742 (99.98%) | 5,948 (21.6%) | |
| 11) | Not worn | | | | < .001 |
| | | Yes | 208 (0.8%) | 23 (11.1%) | |
| | | No | 27,539 (99.3%) | 5,948 (21.6%) | |
| 12) | Not_applicable | | | | < .001 |
| | | Yes | 303 (1.1%) | 38 (12.5%) | |
| | | No | 27,444 (98.9%) | 5,926 (21.6%) | |
| 13) | Unknown (traffic_accident) | | | | 0.12 |
| | | Yes | 5 (0.02%) | 3 (60.0%) | |
| | | No | 27,742 (99.98%) | 5,911 (21.6%) | |
| 14) | Arrival mode | | | | < .001 |
| | | Walk in | 4 (0.0%) | 2 (50.0%) | |
| | | Ambulance in Medical service | 26 (0.1%) | 21 (80.8%) | |
| | | Ambulance in 911 | 7,151 (25.8%) | 1,907 (26.7%) | |
| | | Private Ambulance | 1,137 (4.1%) | 763 (67.1%) | |
| | | Public Mobility | 17 (0.1%) | 3 (17.6%) | |
| | | Other Mobility | 18,368 (66.2%) | 3,233 (17.6%) | |

**TABLE 2.** ***(Continued.)*** **Predictive variables and outcomes for 27,747 ED patients.**

| | | | | | |
|---|---|---|---|---|---|
| 15) | Triage | | | | < .001 |
| | | Level 1 | 471 (1.7%) | 312 (66.2%) | |
| | | Level 2 | 2,372 (8.5%) | 1,302 (54.8%) | |
| | | Level 3 | 14,216 (51.2%) | 3,848 (27.1%) | |
| | | Level 4 | 7,333 (26.4%) | 442 (6.0%) | |
| | | Level 5 | 2,061 (7.4%) | 45 (2.2%) | |
| | | Others | 1,294 (4.7%) | 0 (0.0%) | |
| 16) | Reaction status | | | | < .001 |
| | | Alert | 25,005 (90.1%) | 4,952(19.8%) | |
| | | Verbal response | 1,179 (4.2%) | 667(56.6%) | |
| | | Unresponsive | 190 (0.7%) | 88(46.3%) | |
| | | Painful response | 372 (1.3%) | 242(65.1%) | |
| | | Nothing | 1,001 (3.6%) | 0(0.0%) | |
| 17) | Chief_complain | | | | < .001 |
| | | Abdominal pain | 3,349 (0.121) | 831 (24.8%) | |
| | | Fever | 2,602 (0.094) | 386 (14.8%) | |
| | | Dizziness | 1,381 (0.05) | 199 (14.4%) | |
| | | Dyspnea | 1,123 (0.04) | 422 (37.6%) | |
| | | Headache | 915 (0.033) | 99 (10.8%) | |
| 18) | Vital signs | | Mean (std) | | < .001 |
| | | Systolic_bp | 116.613 (52.177) | 133.5057993 | |
| | | Diastolic_bp | 68.534 (30.945) | 78.25567322 | |
| | | Pulse | 88.620 (30.144) | 95.84619264 | |
| | | Respiration | 20.276 (5.879) | 21.9850395 | |
| | | Temperature | 35.036 (7.892) | 36.84266263 | |
| | | Ox_saturation | 92.4071 (21.884) | 96.55605984 | |
| 19) | Arrival time | | | | < .001 |
| | | 1–6 | 4,017 (14.5%) | 627 (15.6%) | |
| | | 7–12 | 7,136 (25.7%) | 1,830 (25.6%) | |
| | | 13–18 | 8,355 (30.1%) | 2,141 (25.6%) | |
| | | 19–24 | 8,239 (29.7%) | 1,351 (16.4%) | |
| 20) | Cause of injury | | | | .001 |
| | | Struck | 2,170 (35.2%) | 153 (7.1%) | |
| | | Slip | 1,331 (21.6%) | 197 (14.8%) | |
| | | Cut | 711 (11.5%) | 37 (5.2%) | |
| | | Car | 320 (5.2%) | 23 (7.2%) | |
| | | Pedestrian | 303 (4.9%) | 38 (12.5%) | |
| | | Motorcycle | 296 (4.8%) | 39 (13.2%) | |
| | | Fall | 211 (3.4%) | 44 (20.9%) | |
| | | Poisoning | 146 (2.4%) | 32 (21.9%) | |
| | | Fire | 135 (2.2%) | 0 (0.0%) | |
| | | Bicycle | 118 (1.9%) | 10 (8.5%) | |
| | | Machine | 37 (0.6%) | 7 (18.9%) | |
| | | Choking | 19 (0.3%) | 0 (0.0%) | |
| | | Unknown | 18 (0.3%) | 2 (11.1%) | |
| 21) | Injury status | | | | < .001 |
| | | Accidental | 5,696 (92.4%) | 540 (9.5%) | |
| | | Self_harmed | 143 (2.3%) | 34 (23.8%) | |
| | | Assaulted | 319 (5.2%) | 25 (7.8%) | |
| | | Unspecified | 6 (0.1%) | 1 (16.7%) | |
| | | Other | 1 (0.0%) | 0 (0.0%) | |
| | | – | 5,349 (77.8%) | 5,349 (89.9%) | |
| 22) | Hospitalization | | 5,949 (21.4%) | | < .001 |

set of rules to create partitions of the generated data and iterates the partitioning process to produce predictions. DT can classify data without complicated calculations and can be used for both categorical and classification variables. It is generally suitable for predicting categorical outcomes [34].

#### 5) NATURAL GRADIENT BOOSTING (NGBOOST)

NGBoost, proposed by Duan *et al.* [22], is a supervised learning algorithm with stochastic prediction capabilities. It estimates the parameters of the conditional probability distribution $P(y|x)$ as a function of $x$ by boosting. NGBoost outputs the overall probability distribution for predictions using natural gradients [35].

### C. STUDY SETTING

This section describes the fitting process of the predictive models using the hyper-parameter tuning and a prediction evaluation method. We also present two types of experiments for hospitalization predictions and explain the derivation of feature importance. Finally, we show how the quantitative effects of hospitalization predictions could be estimated.

#### 1) MODEL FITTING AND EVALUATION

The predictive models are all tested under the same conditions with a training dataset of 19,422 (70%) and a test dataset of 8,325 (30%) randomly selected samples. When LR- and SVM- based predictive models are tested, recursive variable elimination and cross-validation methods are used to extract variables that maximize performance. The DT, XGBoost, and NGBoost models, which are based on embedded methods, are structured to select the features that contribute to the models' *accuracy*; thus, separate feature selection is not required. The predictive models are optimized with hyper-parameters that maximized *AUC* by 10-fold cross-validation. Using an optimized combination of hyper-parameters, 19,422 (70%) samples are trained, excluding the test set. Subsequently, the performance of the models on the test set is reviewed within a 95% confidence interval.

#### 2) PREDICTION AND PERFORMANCE

For the hospitalization predictions, we conduct the following experiments.

**TABLE 3.** **Confusion matrix.**

| Total number of instances = $N$ | | Predictive class | |
|---|---|---|---|
| | | Discharge from ED | Hospitalization |
| Actual class | Discharge from ED | TN (true negative) | FP (false positive) |
| | Hospitalization | FN (false negative) | TP (true positive) |

##### *a: PERFORMANCE COMPARISON FOR PREDICTION MODELS*

This analysis uses five machine learning algorithms (LR, XGBoost, NGBoost, DT, and SVM) to compare the predictive outcomes for ED patients' hospitalizations. We predict hospitalization for ED patients using 27,747 samples. The predictive results are presented in a confusion matrix, and they are interpreted. Table 3 represents a confusion matrix for the predictive results of this study. A confusion matrix is a concept in machine learning that contains results about actual and predictive classification performed by classification algorithms. The confusion matrix has two dimensions. The actual class of the object indexes one dimension, and the class predicted by the classifier indexes the other dimension [36].

This study postulates that if hospitalization predictions are obtained within 20 min after patients arrive at an ED, it would positively lead to the reduction of ED overcrowding. Therefore, it is desirable to select a model with a high true-positive rate; the true-positive rate (also called recall or *sensitivity*) is defined as the measure of a predictive model's ability to select true positive cases from among actual positive cases. In some cases, patients discharged from the ED are incorrectly predicted to be hospitalized (false positives). Then, outpatient hospitalization is delayed due to the reserved empty inpatient beds for the ED patients, and time is wasted on keeping the beds empty. Consequently, *specificity*, the true-negative rate (i.e., true-negative cases divided by actual negative cases), should also be considered to avoid wasting time in inpatient beds when researchers are selecting the best predictive model. *Accuracy* is the ratio of the total number of correctly predicted predictions, and it increases even if the true-negative rate, which may not be of most importance, increases.

However, it is important to verify the reliability of the predictive model for a new dataset. *AUC* is a comprehensive performance measure for all possible classification thresholds, and it is scale-invariant. It measures the quality of the predictive model, regardless of the classification threshold that is selected. Therefore, in this study, we use *AUC* to select the best predictive model.

#### b: PERFORMANCE COMPARISON FOR MORE TRAINING SAMPLES

One of the critical questions in predictive modeling research is how performance changes according to the amount of training data [10]. To test the potential benefits of using more data in modeling, the size of each dataset in this experiment is changed depending on the number of days. We test predictive performance by gradually increasing the amount of data (10, 20, 30, 60, 90, 120, 150, 180, and 212 days). The size of each dataset is determined by calculating the average number of patients per day from the collected data and multiplying it by the number of days specified for each dataset. Then, each dataset is randomly constructed from the data obtained during the seven months. The experiment show which predictive algorithm is most effective in handling datasets containing less than 27,747 patient data (i.e., data available for this study).

#### c: FEATURE IMPORTANCE

Feature importance is derived from tree-based algorithms. Every node creates a set of similar samples using parameters (Gini index or entropy [37]) that remove impurities for the variables in a DT. In this study, we derive the feature importance of hospitalization predictions using the Gini index for XGBoost and NGBoost predictions.

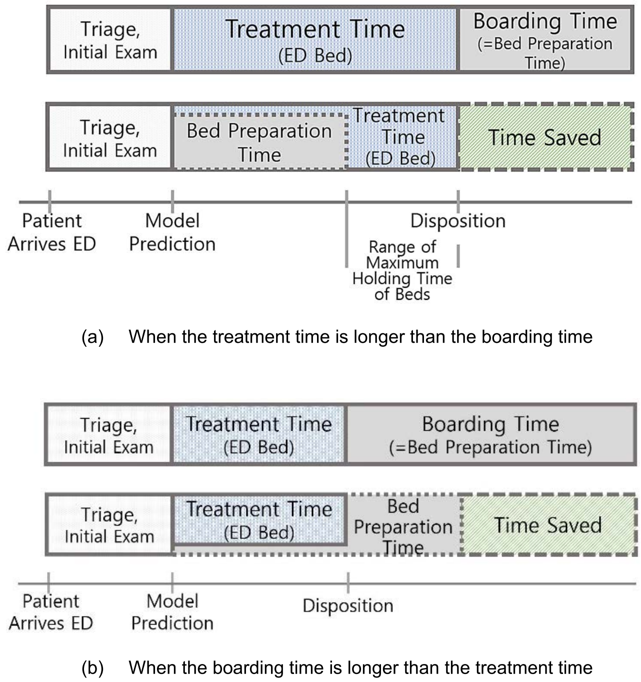



**FIGURE 2.** **Case of patients flow in ED.**

#### d: ESTIMATING TIME EFFECTS OF HOSPITALIZATION PREDICTIONS

This section describes how to estimate the consequences of hospitalization predictions. The estimation method for the consequences of hospitalization predictions reflects patients' flow in the ED and administrative processes. The typical patient flow in the ED is as follows. After a patient enters the ED, classification based on severity is performed at the triage stage. Initial examinations, such as blood and reaction tests, are then completed. All the data used to predict hospitalization in ED patients are acquired through triage, initial examinations, and registration. During treatment, physicians diagnose and treat the patients. After treatment is completed, physicians decide whether the patient needs to be directly discharged from the ED or hospitalized to an inpatient bed; this decision is called *disposition*.

We briefly describe the effects of hospitalization predictions in the ED. This study emphasizes that predictive models can enable decision-making about hospitalization to be made early within 20 minutes of patients entering the ED. The administration can prepare an inpatient bed for the ED patient to be hospitalized in advance. Accordingly, it is possible to shorten the boarding time for the ED patient to wait for the inpatient bed after their treatment. In other words, hospitalization predictions in the ED can positively affect the overcrowding by reducing patients' boarding time and ED length of stay. When the patient's hospitalization is decided, an inpatient bed in a ward is prepared for the patient. We assume that the boarding time is the same as the bed preparation time, and the hospitalization prediction for the patient is available 20 min after entering the ED.

The ED patients' flow is categorized into two cases. First, as shown in Fig. 2(a), there is the case wherein an ED patient whose boarding time is shorter than the treatment time in the ED. If the patient is predicted to be hospitalized after the results of the initial examinations are available, bed preparation for the patient begins in the wards. While the inpatient bed is in preparation, the patient will still be treated in the ED.

If the bed preparation is completed before the treatment ends, the inpatient bed in the wards remains empty until the treatment ends. The prediction effect depends on how long the inpatient bed in the wards can be kept empty. *The maximum holding time of the beds* is defined as the maximum duration of time a bed in a ward reserved for an ED patient is kept empty.

This study observes the time effects of hospitalization predictions according to the varying *maximum holding time of beds*. Second, as shown in Figure. 2(b), if the boarding time is longer than the treatment time, the bed preparation is not completed until the treatment ends. In this case, the patient's LOS in the ED is reduced by the difference between the times at which the original hospitalization decision (i.e., *disposition*) is made and a prediction is made. It is assumed that the inpatient bed for the ED patient is occupied as soon as it is ready.

The quantitative effect of hospitalization predictions depends on the quality of the predictions. When new data are provided, the models we use can predict the probability that a patient is hospitalized. The classification threshold is defined as a reference point that classifies a prediction as either hospitalization or non-hospitalization. If the predicted probability of hospitalization is higher than the classification threshold, the patient is classified as a predicted inpatient. This study estimates the time effects of hospitalization predictions according to prediction quality using the classification threshold. We find classification thresholds that yielded *specificities* of 90, 95, 99, and 99.9%, respectively. Here, we use data from patients who are not hospitalized. According to the obtained classification threshold, all the patients are classified as either predicted inpatients or non-hospitalized patients. Using the classified results, we could estimate the *time saved in the ED* and *the time costs in the inpatient wards* for hospitalization predictions, referring to Algorithms 1 and 2.

We define *the time saved in the ED* as the reduced ED waiting time due to the ED patients' hospitalization predictions and the *time costs in the inpatient ward* as the time is taken to maintain an empty bed in the ward. Table 4 explains the variables included in Algorithms 1 and 2.

**TABLE 4. Notations for the pseudocode.**

| Terms | Description |
|---|---|
| time_saved_in_ED | Reduced ED waiting time due to the ED patients' hospitalization predictions |
| disposition | Time at which whether a patient is hospitalized or directly discharged from ED is decided |
| model_prediction | Time at which a patient's hospitalization is forecasted |
| boarding_time | Time between the hospitalization decision and departure from the ED [12] |
| bed_preparation _complete | Time at which inpatient bed preparation for ED patients is completed |
| max_holding | Maximum amount of time an empty bed is held for a patient who is predicted to be hospitalized (*Maximum holding time of beds*) |
| TP_bed_ready_time | Time at which inpatient bed preparation for a true-positive patient is completed |
| TP_time_cost | Time between the completion of inpatient bed preparation for a true-positive patient and the occupancy of the inpatient bed by the patient |
| FP_bed_ready_time | Time at which inpatient bed preparation for a false-positive patient is completed |
| avg_boarding_time | Average boarding time |
| non_inpatient_LOS | LOS of ED patients who are not hospitalized |
| FP_time_cost | Time between the completion of inpatient bed preparation for a false-positive patient and the discharge of the patient from the ED |

## IV. RESULTS

### A. PERFORMANCE COMPARISON FOR PREDICTION MODULES

This section describes the results of the hospitalization predictions for a test dataset (8,325 patients). All five hospitalization predictive models show high discernment ability. In Table 5, for 1,788 hospitalized patients, SVM predicts 1,248 (69.8%) patients, and XGBoost predicts 1,196 (66.9%) patients to be hospitalized. In contrast, for 6,510 non-hospitalized patients, NGBoost most accurately predicts 6,352 (97.6%) patients to be non-hospitalized.

The AUC, accuracy, sensitivity, and specificity of each model are summarized in Table 6. Table 6 summarizes that SVM has an accuracy of 0.8961 (95% CI 0.88–0.90), and XGBoost also has an accuracy of 0.8961 (95% CI 0.64–0.91). SVM and XGBoost display the highest accuracy.

Although NGBoost has the lowest accuracy, it has the highest specificity of 0.9717 (95% CI 0.97–0.98). XGBoost has the second-highest specificity of 0.9582 (95% CI 0.58–0.98). XGBoost has the highest AUC of 0.9332 (95% CI 0.92–0.94).

### B. PERFORMANCE COMPARISON FOR MORE TRAINING SAMPLES

Table 7 lists the size of each dataset according to the collection period. In Table 7, up to 30 days, we use 10-day increments. It is expected that a significant performance gain can be achieved if data are added for predictions, when datasets used for predictions are relatively small. We present the results of this experiment in Table 8. Table 8 summarizes *the AUC* of each predictive model depending on the size of

**Algorithm 1** Estimating *Time Saved in the ED* for an ED Patient

```
Time_saved_in_ED is set as the difference between disposition and model_prediction
if boarding_time < time_saved_in_ED
    time_saved_in_ED = boarding_time
Else
    time_saved_in_ED = time_saved_in_ED
    if disposition – bed_preparation_complete ≤ max_holding
        time_saved_in_ED = time_saved_in_ED
    else
        time_saved_in_ED = 0
end
```

**Algorithm 2** Estimating *Time Costs in Inpatient Beds* for an ED Patient

```
TP_bed_ready_time is set as the sum of model_prediction and boarding_time.
# time costs in case of true positive
if disposition – TP_bed_ready_time > 0
    TP_time_cost = disposition – TP_bed_ready_time
else
    TP_time_cost = 0
TP_time_cost is redefined as the minimum value between max_holding and TP_time_cost.
# time costs in case of false positive
FP_bed_ready_time is set as the sum of model_prediction and avg_boarding_time.
if non_inpatient_LOS – FP_bed_ready_time > 0
    FP_time_cost = non_inpatient_LOS – FP_bed_ready_time
else
    FP_time_cost = 0
FP_time_cost is redefined as the minimum value between max_holding and FP_time_cost.
end
```

the dataset. As the size of the dataset increases, *the AUC* of each model improves. While the *AUC* of LR, XGBoost, and SVM slightly increases (i.e., less than 1%) as the size of the dataset increases, *the AUC* of DT increases by approximately 3% or more as the dataset increases in size from the 10days dataset to the 60 days dataset.

In other words, *the AUC* of the DT is remarkably degraded for small datasets. *The AUC* of XGBoost is the highest regardless of the dataset size. LR and NGBoost show relatively high *AUC* values when comparing the dataset collected after 10 days to that collected after 120 days. LR,

NGBoost, and SVM show similar *AUC* values for the 150 days dataset. Fig. 3 shows the *AUC* of each model according to the size of the dataset. Fig. 4 shows the *sensitivity* of the three models that shows a higher *AUC* than the other models. SVM has the highest *sensitivity* for the entire test dataset (Table 6); however, it had a significantly low sensitivity for small datasets (i.e., 10 and 20-day datasets).

To focus on true-positive predictions, SVM can be selected for datasets with a size of 180 days or more. However, it is not recommended to select it for datasets with a size less than or equal to 30 days.

**TABLE 5. Hospitalization predictions results represented as confusion matrices.**

| LR, $N = 8,325$ | Predicted discharge from ED | Predicted hospitalization |
|---|---|---|
| Actual discharge from ED | 6,241 | 296 |
| Actual predicted hospitalization | 622 | 1,166 |

(c) Confusion matrix for LR based predictive model results

| XGB, $N = 8,325$ | Predicted discharge from ED | Predicted hospitalization |
|---|---|---|
| Actual discharge from ED | 6,264 | 273 |
| Actual predicted hospitalization | 592 | 1,196 |

(d) Confusion matrix for XGBoost based predictive model results

| NGBoost, $N = 8,325$ | Predicted discharge from ED | Predicted hospitalization |
|---|---|---|
| Actual discharge from ED | 6,352 | 185 |
| Actual predicted hospitalization | 972 | 816 |

(e) Confusion matrix for NGBoost based predictive model results

| SVM, $N = 8,325$ | Predicted discharge from ED | Predicted hospitalization |
|---|---|---|
| Actual discharge from ED | 6,212 | 325 |
| Actual predicted hospitalization | 540 | 1,248 |

(f) Confusion matrix for SVM based predictive model results

| DT, $N = 8,325$ | Predicted discharge from ED | Predicted hospitalization |
|---|---|---|
| Actual discharge from ED | 6,071 | 466 |
| Actual predicted hospitalization | 618 | 1,170 |

(g) Confusion matrix for DT based predictive model results

**TABLE 6. Hospitalization prediction results based on performance indicators.**

| Algorithm | Test *AUC* (95% CI) | Test *Accuracy* (95% CI) | Test *Sensitivity* (95% CI) | Test *Specificity* (95% CI) |
|---|---|---|---|---|
| LR | 0.9143 (0.90–0.92) | 0.8885 (0.88–0.90) | 0.6521 (0.61–0.68) | 0.9547 (0.94–0.96) |
| XGBoost | 0.9332 (0.92–0.94) | 0.8961 (0.64–0.90) | 0.6689 (0.31–0.92) | 0.9582 (0.58–0.98) |
| NGBoost | 0.9137 (0.90–0.92) | 0.8610 (0.85–0.87) | 0.4564 (0.42–0.48) | 0.9717 (0.97–0.98) |
| SVM | 0.9074 (0.89–0.91) | 0.8961 (0.88–0.90) | 0.6980 (0.67–0.71) | 0.9503 (0.94–0.96) |
| DT | 0.8754 (0.86–0.89) | 0.8698 (0.84–0.88) | 0.6544 (0.49–0.72) | 0.9287 (0.91–0.95) |

**TABLE 7.** Size of the dataset according to collection period.

| Data set | 10 days | 20 days | 30 days | 60 days | 90 days | 120 days | 150 days | 180 days | 212 days |
|---|---|---|---|---|---|---|---|---|---|
| Train (70%) | 948 | 1,801 | 2,660 | 5,394 | 8,352 | 11,016 | 13,690 | 16,325 | 19,422 |
| Test (30%) | 407 | 773 | 1,140 | 2,312 | 3,580 | 4,722 | 5,868 | 6,997 | 8,325 |
| Total | 1,355 | 2,574 | 3,800 | 7,706 | 11,932 | 15,738 | 19,558 | 23,322 | 27,747 |

**TABLE 8.** AUC of each model according to data size.

| Days | LR (95% CI) | XGBoost (95% CI) | NGBoost (95% CI) | SVM (95% CI) | DT (95% CI) |
|---|---|---|---|---|---|
| 10 | 0.8671 (0.83–0.95) | 0.8739 (0.83–0.96) | 0.8378 (0.79–0.94) | 0.8331 (0.75–0.94) | 0.7120 (0.65–0.84) |
| 20 | 0.8707 (0.87–0.92) | 0.8816 (0.87–0.92) | 0.8716 (0.84–0.92) | 0.8335 (0. 82–0.90) | 0.7674 (0.76–0.84) |
| 30 | 0.8822 (0.87–0.93) | 0.9031 (0.90–0.95) | 0.8754 (0.85–0.93) | 0.8484 (0.84–0.92) | 0.8066 (0.77–0.86) |
| 60 | 0.8865 (0.88–0.92) | 0.9161 (0.90–0.94) | 0.8888 (0.86–0.91) | 0.8674 (0.84–0.90) | 0.8293 (0.82–0.88) |
| 90 | 0.8933 (0.89–0.93) | 0.9157 (0.91–94) | 0.8898 (0.86–0.90) | 0.8771 (0.87–0.92) | 0.8524 (0.84–0.89) |
| 120 | 0.8996 (0.90–0.92) | 0.9229 (0.91–94) | 0.8921 (0.87–0.91) | 0.8781 (0.85–0.89) | 0.8633 (0.86–0.89) |
| 150 | 0.9073 (0.90–0.92) | 0.9291 (0.92–0.94) | 0.8923 (0.87–0.89) | 0.9065 (0.89–0.92) | 0.8685 (0.86–0.88) |
| 180 | 0.9076 (0.90–0.92) | 0.9296 (0.91–0.95) | 0.9029 (0.88–0.91) | 0.9074 (0.90–0.92) | 0.8696 (0.86–0.89) |
| 210 | 0.9143 (0.90–0.92) | 0.9332 (0.92–0.94) | 0.9137 (0.90–0.92) | 0.9074 (0.89–0.91) | 0.8754 (0.86–0.89) |

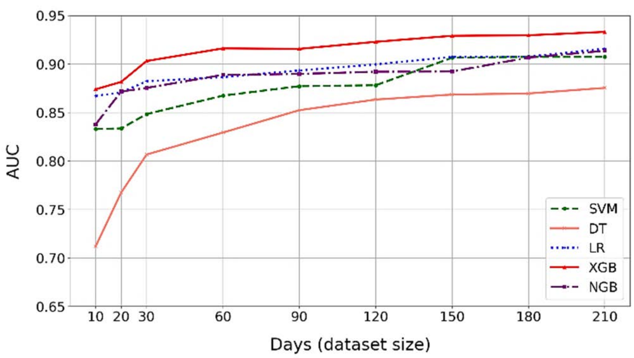



**FIGURE 3.** ***AUC*** **comparison between more training samples.**

### C. FEATURE IMPORTANCE

Feature importance can be derived from tree-based hospitalization predictive models. This study compares the results of feature importance from XGBoost (see Fig. 5 (a)) and NGBoost (see Fig. 5 (b)). The two models are selected for having the highest *AUC* values among the tree-based models.

XGBoost shows high information gain from the triage level, reaction status, vital signs, demographics, and main

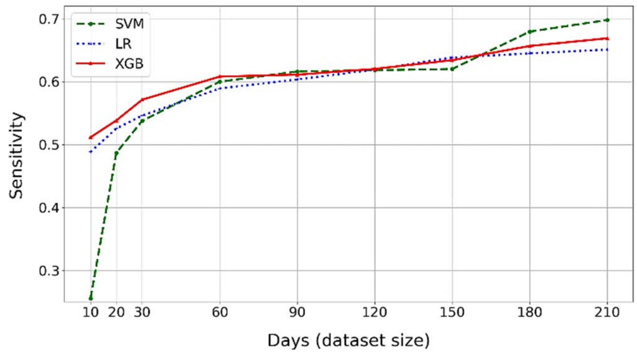
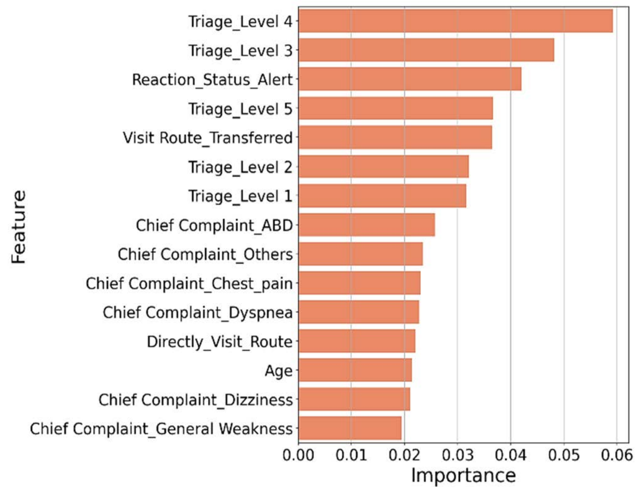
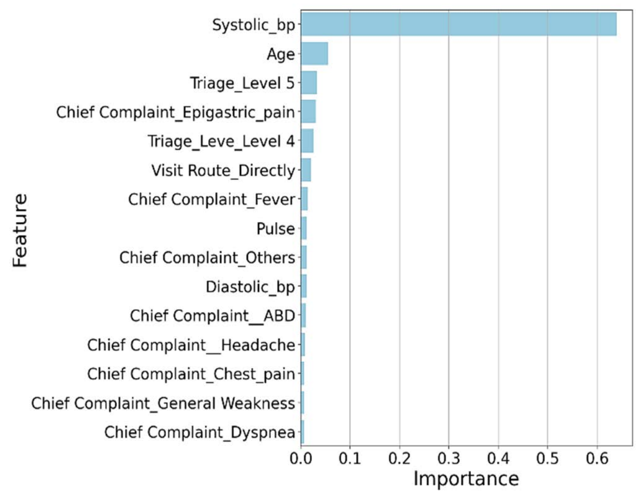



**FIGURE 4.** ***Sensitivity*** **comparison between more training samples.**




(a) Feature importance from XGBoost




(b) Feature importance from NGBoost.

**FIGURE 5.** Feature importance from XGBoost and NGBoost.

complaints of the patients. For patients with triage levels 3, 4, and 5, the probability of hospitalization is low; thus, these variables may have contributed to the prediction of

**TABLE 9.** Time effects of XGBoost hospitalization predictions.

min

| Specificity | Maximum holding time of beds | ED time saved for 8,325 patients | Bed time costs for 8,325 patients | ED time saved for a year | Bed time costs for a year | ED time saved for an ED patient | Bed time costs for an ED patient |
|---|---|---|---|---|---|---|---|
| 90% | 30 | 1,856.000 | 39,579.598 | 6,185.998 | 131,917.730 | 0.223 | 4.754 |
| | 60 | 17,362.000 | 74,233.635 | 57,867.077 | 247,418.699 | 2.086 | 8.917 |
| | 120 | 61,168.000 | 130,574.745 | 203,871.291 | 435,202.098 | 7.348 | 15.685 |
| | 240 | 118,734.000 | 197,605.929 | 395,737.213 | 658,615.222 | 14.262 | 23.736 |
| 95% | 30 | 1,557.000 | 29,698.122 | 5,189.439 | 98,983.039 | **0.19** | **3.567** |
| | 60 | 15,048.000 | 55,818.588 | 50,154.577 | 186,041.846 | **1.808** | **6.705** |
| | 120 | 51,465.000 | 98,821.005 | 171,531.454 | 329,367.738 | **6.182** | **11.870** |
| | 240 | 102,055.000 | 148,948.482 | 340,146.557 | 496,441.266 | **12.259** | **17.892** |
| 99% | 30 | 868.000 | 11,846.273 | 2,893.021 | 39,483.307 | 0.104 | 1.423 |
| | 60 | 7,016.000 | 22,461.273 | 23,384.138 | 74,862.815 | 0.843 | 2.698 |
| | 120 | 22,801.000 | 39,500.841 | 75,995.117 | 131,655.235 | 2.739 | 4.745 |
| | 240 | 48,964.000 | 57,635.364 | 163,195.689 | 192,097.109 | 5.882 | 6.923 |
| 99.90% | 30 | 307.000 | 3,530.400 | 1,023.223 | 11,766.728 | 0.037 | 0.424 |
| | 60 | 2,881.000 | 6,606.400 | 9,602.295 | 22,018.953 | 0.346 | 0.794 |
| | 120 | 7,675.000 | 11,646.800 | 25,580.568 | 38,818.470 | 0.922 | 1.399 |
| | 240 | 15,541.000 | 16,695.600 | 51,797.733 | 55,645.984 | 1.867 | 2.005 |

non-hospitalized patients. NGBoost has the highest *specificity* (i.e., the highest ratio of true-negative predictions). Therefore, most of the variables in the NGBoost feature importance may relate more to non-hospitalized predictions. Appendices A and B show the feature importance of XGBoost and NGBoost according to the size of the dataset.

### D. ESTIMATING TIME EFFECTS OF HOSPITALIZATION PREDICTIONS

Table 9 shows the estimated time effects of the hospitalization predictions. To estimate the time effects, we use the predictive results of XGBoost based on all the performance indicators (i.e., *AUC, accuracy, sensitivity*, and *specificity*), as shown in Table 6. The *time saved in the ED* and time costs in inpatient beds are calculated while adjusting the *specificity* from 90 to 99.9% and *the maximum holding time of beds* from 30 to 240 min.

When the *maximum holding time of beds* is 240 min, the *time saved in the ED* per patient is estimated to be 12.3 min at 95% specificity, which is close to the *specificity* (95.82%) of XGBoost. *The time costs of inpatient beds* per patient are estimated to be 17.9 min.

According to the predictions of the five predictive models at 95% specificity, the quantitative effects of hospitalization predictions are compared. The most reliable results are the time effects obtained using the predictive results of the best-performing model. In this study, since XGBoost is the best performance model, the predictive results are regarded as the most reliable among the five predictive models. The results of XGBoost show higher ED time saved and inpatient bed time costs for an ED patient (see Table 10).

## V. DISCUSSION

We construct and test LR, XGBoost, NGBoost, SVM, and DT models to predict hospitalization for ED patients. This study shows that hospitalization can be predicted precisely (at a high AUC from 0.89 to 0.92) using the information that is available within 20 min after a patient enters the ED. Accurate hospitalization predictions can potentially reduce ED overcrowding, improve the quality of patient care, and support the implementation of proper treatment resources [12].

We train five predictive models by varying the size of the dataset and examine whether the performance of the models improves when more data are used for hospitalization predictions. The results show that as more data are used for predictions, the *AUC* values of the predictive models increase. Although this study uses a small amount of data compared to other studies, it shows that the predictive models trained on a relatively small dataset (i.e., larger than or equal to a 150days dataset) provide hospitalization predictions at high *AUC* values (from 0.8754 to 0.9332). In particular, the predictive models perform well for all ED patients without any restrictions based on the analysis target (i.e., patients with a specific disease or patients of a specific age group).

Duan *et al.* [22] hypothesized that the NGBoost model used in our experiments would perform well, even with a small-size dataset. Although hyper-parameter searching and tuning are done for NGBoost, *the accuracy* of NGBoost (0.8610)

**TABLE 10.** Time effects of all hospitalization prediction at 95% specificity.

min

| Algorithms | Maximum holding time of beds | ED time saved for 8,325 patients | Bed time costs for 8,325 patients | ED time saved for a year | Bed time costs for a year | ED time saved for an ED patient | Bed time costs for an ED patient |
|---|---|---|---|---|---|---|---|
| XGB | 30 | 1,557.000 | 29,698.122 | 5,189.439 | 98,983.039 | **0.19** | **3.567** |
| | 60 | 15,048.000 | 55,818.588 | 50,154.577 | 186,041.846 | **1.808** | **6.705** |
| | 120 | 51,465.000 | 98,821.005 | 171,531.454 | 329,367.738 | **6.182** | **11.870** |
| | 240 | 102,055.000 | 148,948.482 | 340,146.557 | 496,441.266 | **12.259** | **17.892** |
| LR | 30 | 1,360.000 | 28,000.189 | 4,532.843 | 93,323.874 | **0.163** | **3.363** |
| | 60 | 12,924.000 | 52,432.051 | 43,075.343 | 174,754.609 | **1.552** | **6.298** |
| | 120 | 46,928.000 | 92,701.567 | 156,409.756 | 308,971.819 | **5.637** | **11.135** |
| | 240 | 96,266.000 | 138,817.101 | 320,851.976 | 462,673.646 | **11.563** | **16.675** |
| NGB | 30 | 1,330.000 | 25,114.433 | 4,432.854 | 83,705.727 | **0.160** | **3.017** |
| | 60 | 12,638.000 | 47,245.637 | 42,122.112 | 157,468.430 | **1.518** | **5.675** |
| | 120 | 41,740.000 | 82,957.205 | 139,118.292 | 276,494.123 | **5.014** | **9.965** |
| | 240 | 84,053.000 | 124,285.153 | 280,146.377 | 414,239.054 | **10.096** | **14.929** |
| DT | 30 | 1,241.000 | 24,984.244 | 4,136.219 | 83,271.810 | **0.149** | **3.001** |
| | 60 | 12,140.000 | 46,922.026 | 40,462.292 | 156,389.846 | **1.458** | **5.636** |
| | 120 | 40,716.000 | 82,148.335 | 135,705.328 | 273,798.179 | **4.891** | **9.868** |
| | 240 | 80,087.000 | 122,780.720 | 266,927.806 | 409,224.821 | **9.620** | **14.748** |
| SVM | 30 | 1,567.000 | 29,605.375 | 5,222.769 | 98,673.915 | **0.19** | **3.556** |
| | 60 | 14,918.000 | 55463.923 | 49,721.291 | 184,859.755 | **1.792** | **6.662** |
| | 120 | 50,665.000 | 98,055.845 | 168,865.076 | 326,817.482 | **6.086** | **11.778** |
| | 240 | 101,183.000 | 147,453.509 | 337,240.204 | 491,458.561 | **12.154** | **17.712** |

when working on the entire dataset is worse than that of the other models (0.6521 to 0.6980). Therefore, we conclude that NGBoost is not appropriate in improving the proportion of correct hospitalization predictions; however, NGBoost has high *specificity*. In this study, 21.4% of the ED patients are hospitalized (i.e., minor class), and 78.6% are non-hospitalized (i.e., major class). NGBoost has the ability to predict a major class. Therefore, this study demonstrates the possibility of deploying NGBoost to predict a major class in the ED (e.g., triage predictions for level 3).

The quantitative effects of hospitalization predictions are estimated as the *time saved in the ED* and *time costs in inpatient beds*. For these estimations, we use the predictive results of XGBoost. When the *maximum holding time of the beds* is set to 240 min, the LOS for an ED patient is reduced by 12.3 min (i.e., *time saved in the ED* for a patient) and that for all patients per year is reduced by 340,147 min, as shown in Table 9. These results indicate that hospitalization predictions help reduce ED overcrowding. However, the time requires to hold an empty inpatient bed for ED patients is 17.9 min (i.e., *time costs in inpatient beds* for an ED patient), and the total time to hold empty inpatient beds for all ED patients per year is 496,441 min. As shown in Table 9, *the time costs in inpatient beds* are approximately 1.5 times higher than the *time saved in the ED*. However, *time costs in inpatient beds* are calculated under the postulation that holding an empty inpatient bed always prevents the inpatient bed occupancy of an incoming patient.

**TABLE 11.** Average inpatient bed utilization of local hospitals in South Korea (national medical center in South Korea).

| Year | '16 | '17 | '18 | '19 |
|---|---|---|---|---|
| Inpatient bed utilization | 87.68 | 88.04 | 81.58 | 83.37 |

* formula: number of inpatients / (number of permitted beds×365) × 100

The quantitative effects of hospitalization predictions can be interpreted according to three perspectives. First, the significant proportion of *time costs in inpatient beds* estimated in this study may not be the actual cost spent on preventing the inpatient bed occupancy of incoming patients. As shown in Table 11, the average inpatient bed utilization at local medical centers in South Korea is less than 90% [38]. In addition, the influx of outpatients for hospitalization is scarce during the night. ED patients do not always compete with outpatients for inpatient beds. Second, *time costs for inpatient beds* are indeed less significant than their nominal value, considering the number of inpatient beds. The number of inpatient beds at KUAH is 1,100, and *time costs in inpatient beds* are 17.9 min for an ED patient when the *maximum holding time* of the beds is 240 min. In contrast, the number of ED beds is 35 and *the time saved in the ED* is 12.3 min for an ED patient. The 17.9 min *time costs in inpatient beds* are dispersed across 1,100 inpatient beds. Finally, given that the LOS in inpatient beds is much longer than that in the ED, *the time costs in inpatient beds* are less significant than their nominal values. At KUAH, the average LOS in the ED is approximately 420 min, whereas the average LOS in inpatient beds is approximately 6.8 days. The *time costs in inpatient beds* of 17.9 min constitute a very small portion of the average LOS in inpatient beds. It is concluded that *the time saved in the ED* (i.e., 12.3 min for an ED patient) is significant considering the small number of ED beds and the short LOS, whereas *time costs in inpatient beds* (i.e., 17.9 min for an ED patient) may not be high considering a large number of inpatient beds and the long LOS of patients in inpatient beds.

## VI. CONCLUSION

This study shows that predictive models could provide quality predictions for hospitalization using primary information on patients that could be obtained within 20 min of their ED entrance. XGBoost provides hospitalization predictions with the best *AUC* (0.9332). LR, NGBoost, SVM, and DT predictions generally show good *AUC* values (0.8754–0.9143). NGBoost, which is used for the first time to predict hospitalization, might not be appropriate for predicting a minor class.

By adjusting *the sensitivity* of the XGBoost model and the *maximum holding time of beds* (administrative policy to be followed), the time effects of hospitalization predictions are estimated. The results show that hospitalization predictions could be utilized to reduce patients' LOS in the ED. Therefore, we expect that accurate hospitalization predictions will alleviate ED overcrowding.

In this study, using a relatively small amount of data for predictions could be viewed as both a distinction and limitation. The experiments in this study do not show convergence in *AUC* values and *sensitivity* with respect to the size of the dataset. In particular, *sensitivity* is significantly lower than that of other indicators because of the small number of hospitalized ED patients. We may use more data in further studies to observe the convergence of *AUC* and *sensitivity*.

This study shows that hospitalization predictions can reduce patients' ED length of stay by shortening decision-making time. These results suggest the possibility of improving ED overcrowding. Apart from the ED overcrowding issue, this study can be extended to investigate how the reduction of the ED length of stay affects the improvement of the patients' health conditions. In addition, it is possible to investigate the effects of shortening the ED length of stay on the number of days hospitalized in inpatient wards.

## APPENDIX A

**TABLE 12.** Feature importance of XGBoost model according to data size.

| Data size | Rank 1 | Rank 2 | Rank 3 | Rank 4 | Rank 5 |
|---|---|---|---|---|---|
| 10 Days | Transferred (Visit route) | Respiration (Vital sign) | Directly (Visit route) | Level 4 (Triage) | Abdominal (Chief complaint) |
| 20 Days | Level 2 (Triage) | Transferred (Visit route) | Directly (Visit route) | Disease (Disease status) | Level 4 (Triage) |
| 30 Days | Transferred (Visit route) | Respiration (Vital sign) | Level 4 (Triage) | Level 2 (Triage) | Directly (Visit route) |
| 60 Days | Level 2 (Triage) | Level 2 (Triage) | Level 2 (Triage) | Level 2 (Triage) | Private ambulance (Arrival mode) |
| 90 Days | Transferred (Visit route) | Level 2 (Triage) | Level 3 (Triage) | Level 4 (Triage) | Level 5 (Triage) |
| 120 Days | Level 3 (Triage) | Directly (Visit route) | Level 4 (Triage) | Level 1 (Triage) | Alert (Reaction status) |
| 150 Days | Level 3 (Triage) | Level 1 (Triage) | Alert (Reaction status) | Level 4 (Triage) | Chest pain (Chief complaint) |
| 180 Days | Transferred (Visit route) | Level 4 (Triage) | Level 5 (Triage) | Level 1 (Triage) | Level 2 (Triage) |
| 212 Days | Level 4 (Triage) | Level 3 (Triage) | Alert (Reaction status) | Level 5 (Triage) | Transferred (Visit route) |

## APPENDIX B

**TABLE 13.** Feature importance of NGBoost model according to data size.

| Data size | Rank 1 | Rank 2 | Rank 3 | Rank 4 | Rank 5 |
|---|---|---|---|---|---|
| 10 Days | Systolic bp (Vital sign) | Age | Directly (Visit route) | Level 4 (Triage) | Respiration (Vital signs) |
| 20 Days | Diastolic bp (Vital sign) | Systolic bp (Vital sign) | Level 5 (Triage) | Age | Level 4 (Triage) |
| 30 Days | Diastolic bp (Vital sign) | Systolic bp (Vital sign) | Headache (Chief complaint) | Epigastric pain (Chief complaint) | Age |
| 60 Days | Systolic bp (Vital sign) | Diastolic bp (Vital sign) | Age | Level 4 (Triage) | Level 5 (Triage) |
| 90 Days | Systolic bp (Vital sign) | Diastolic bp (Vital sign) | Level 5 (Triage) | Age | Level 4 (Triage) |
| 120 Days | Diastolic bp (Vital sign) | Level 5 (Triage) | Age | Directly (Visit route) | Level 4 (Triage) |
| 150 Days | Diastolic bp (Vital sign) | Age | Level 4 (Triage) | Level 5 (Triage) | Directly (Visit route) |
| 180 Days | Diastolic bp (Vital sign) | Epigastric pain (Chief complaint) | Age | Level 5 (Triage) | Level 4 (Triage) |
| 212 Days | Systolic bp (Vital sign) | Age | Level 5 (Triage) | Epigastric pain (Chief complaint) | Level 4 (Triage) |

## ACKNOWLEDGMENT

This study was approved by the Institutional Review Board (IRB) of Korea University Medical Center under

Approval No. 2019AN0531. The authors sincerely thank the Associate Editor and three anonymous reviewers for their constructive comments and suggestions to improve this article.

## REFERENCES


[1] H. M. Jung, M. J. Kim, J. H. Kim, Y. S. Park, H. S. Chung, S. P. Chung, and J. H. Lee, ''The effect of overcrowding in emergency departments on the admission rate according to the emergency triage level,'' *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0247042.

[2] D. B. Chalfin, S. Trzeciak, A. Likourezos, B. M. Baumann, and R. P. Dellinger, ''Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit,'' *Critical Care Med.*, vol. 35, no. 6, pp. 1477–1483, Jun. 2007.

[3] S. L. Bernstein, D. Aronsky, R. Duseja, S. Epstein, D. Handel, U. Hwang, M. McCarthy, K. J. McConnell, J. M. Pines, N. Rathlev, R. Schafermeyer, F. Zwemer, M. Schull, and B. R. Asplin, ''The effect of emergency department crowding on clinically oriented outcomes,'' *Academic Emergency Med.*, vol. 16, no. 1, pp. 1–10, Jan. 2009.

[4] J. S. Peck, J. C. Benneyan, D. J. Nightingale, and S. A. Gaehde, ''Predicting emergency department inpatient admissions to improve same-day patient flow,'' *Acad. Emergency Med.*, vol. 19, no. 9, pp. E1045–E1054, Sep. 2012.

[5] R. M. Cowan and S. Trzeciak, ''Clinical review: Emergency department overcrowding and the potential impact on the critically ill,'' *Crit. Care*, vol. 9, no. 3, pp. 1–5, Jun. 2004.

[6] D. R. Eitel, S. E. Rudkin, M. A. Malvehy, J. P. Killeen, and J. M. Pines, ''Improving service quality by understanding emergency department flow: A white paper and position statement prepared for the American academy of emergency medicine,'' *J. Emergency Med.*, vol. 38, no. 1, pp. 70–79, Jan. 2010.

[7] Y. Barak-Corren, A. M. Fine, and B. Y. Reis, ''Early prediction model of patient hospitalization from the pediatric emergency department,'' *Pediatrics*, vol. 139, no. 5, May 2017, Art. no. e20162785.

[8] J. Wang, J. Li, and P. K. Howard, ''A system model of work flow in the patient room of hospital emergency department,'' *Health Care Manage. Sci.*, vol. 16, no. 4, pp. 341–351, Apr. 2013.

[9] E. W. Dickson, S. Singh, D. S. Cheung, C. C. Wyatt, and A. S. Nugent, ''Application of lean manufacturing techniques in the emergency department,'' *J. Emergency Med.*, vol. 37, no. 2, pp. 177–182, Aug. 2009.

[10] W. S. Hong, A. D. Haimovich, and R. A. Taylor, ''Predicting hospital admission at emergency department triage using machine learning,'' *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0201016.

[11] B. Graham, R. Bond, M. Quinn, and M. Mulvenna, ''Using data mining to predict hospital admissions from the emergency department,'' *IEEE Access*, vol. 6, pp. 10458–10469, 2018.

[12] S. W. Kim, J. Y. Li, P. Hakendorf, D. J. Teubner, D. I. Ben-Tovim, and C. H. Thompson, ''Predicting admission of patients by their presentation to the emergency department,'' *Emergency Med. Australasia*, vol. 26, no. 4, pp. 361–367, Aug. 2014.

[13] J. A. Lucke, J. de Gelder, F. Clarijs, C. Heringhaus, A. J. M. de Craen, A. J. Fogteloo, G. J. Blauw, B. D. Groot, and S. P. Mooijaart, ''Early prediction of hospital admission for emergency department patients: A comparison between patients younger or older than 70 years,'' *Emergency Med. J.*, vol. 35, no. 1, pp. 18–27, Jan. 2018.

[14] O. M. Araz, D. Olson, and A. Ramirez-Nafarrate, ''Predictive analytics for hospital admissions from the emergency department using triage information,'' *Int. J. Prod. Econ.*, vol. 208, pp. 199–207, Feb. 2019.

[15] D. Golmohammadi, ''Predicting hospital admissions to reduce emergency department boarding,'' *Int. J. Prod. Econ.*, vol. 182, pp. 535–544, Dec. 2016.

[16] D. Travers, S. Wu, M. Scholer, M. Westlake, A. Waller, and A. L. McCalla, ''Evaluation of a chief complaint pre-processor for biosurveillance,'' in *Proc. AMIA Annu. Symp.*, 2007, p. 736.

[17] M. M. Dinh, S. B. Russell, K. J. Bein, K. Rogers, D. Muscatello, R. Paoloni, J. Hayman, D. R. Chalkley, and R. Ivers, ''The Sydney triage to admission risk tool (START) to predict emergency department disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia,'' *BMC Emergency Med.*, vol. 16, no. 1, pp. 1–7, Dec. 2016.

[18] A. Fenn, C. Davis, D. M. Buckland, N. Kapadia, M. Nichols, M. Gao, W. Knechtle, S. Balu, M. Sendak, and B. J. Theiling, ''Development and validation of machine learning models to predict admission from emergency department to inpatient and intensive care units,'' *Ann. Emergency Med.*, vol. 78, no. 2, pp. 290–302, Aug. 2021.

[19] T. Goto, C. A. Camargo, M. K. Faridi, R. J. Freishtat, and K. Hasegawa, ''Machine learning–based prediction of clinical outcomes for children during emergency department triage,'' *JAMA Netw. Open*, vol. 2, no. 1, Jan. 2019, Art. no. e186937.

[20] S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, and L. A. Nathanson, ''Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning,'' *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174708.

[21] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, ''Predicting asthma-related emergency department visits using big data,'' *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1216–1223, Jul. 2015.

[22] T. Duan, A. Anand, D. Y. Ding, K. K. Thai, S. Basu, A. Ng, and A. Schuler, ''NGBoost: Natural gradient boosting for probabilistic prediction,'' in *Proc. ICML*, 2020, pp. 2690–2700.

[23] S. Dutta and S. K. Bandyopadhyay, ''Revealing brain tumor using cross-validated NGBoost classier,'' *Int. J. Mach. Learn. Netw. Collaborative Eng.*, vol. 4, pp. 12–20, Jul. 2020.

[24] M. Pfau, S. Sahu, R. A. Rupnow, K. Romond, D. Millet, F. G. Holz, S. Schmitz-Valckenberg, M. Fleckenstein, J. I. Lim, L. de Sisternes, T. Leng, D. L. Rubin, and J. A. Hallak, ''Probabilistic forecasting of anti-VEGF treatment frequency in neovascular age-related macular degeneration,'' *Transl. Vis. Sci. Technol.*, vol. 10, no. 7, p. 30, Jun. 2021.

[25] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. A. Elwafa, and H. Kurdi, ''Impact of dataset size on classification performance: An empirical evaluation in the medical domain,'' *Appl. Sci.*, vol. 11, no. 2, p. 796, Jan. 2021.

[26] F. Chen, Y. Tang, C. Wang, J. Huang, C. Huang, D. Xie, T. Wang, and C. Zhao, ''Medical cyber-physical systems: A solution to smart health and the state of the art,'' *IEEE Trans. Computat. Social Syst.*, early access, Nov. 13, 2021, doi: 10.1109/TCSS.2021.3122807.

[27] K. Potdar, T. S. Pardawala, and C. D. Pai, ''A comparative study of categorical variable encoding techniques for neural network classifiers,'' *Int. J. Comput. Appl.*, vol. 175, no. 4, pp. 7–9, Oct. 2017.

[28] J. W. Yan, K. M. Gushulak, M. P. Columbus, K. van Aarsen, A. L. Hamelin, G. A. Wells, and I. G. Stiell, ''Risk factors for recurrent emergency department visits for hyperglycemia in patients with diabetes mellitus,'' *Int. J. Emergency Med.*, vol. 10, no. 1, pp. 1–8, Dec. 2017.

[29] S. He, B. Li, H. Peng, J. Xin, and E. Zhang, ''An effective cost-sensitive XGBoost method for malicious URLs detection in imbalanced dataset,'' *IEEE Access*, vol. 9, pp. 93089–93096, 2021.

[30] J. C. Stoltzfus, ''Logistic regression: A brief primer,'' *Acad. Emergency Med.*, vol. 18, no. 10, pp. 1099–1104, Oct. 2011, doi: 10.1111/j.1553-2712.2011.01185.x.

[31] T. Chen and C. Guestrin, ''XGBoost: A scalable tree boosting system,'' in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[32] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer Series in Statistics), vol. 1, no. 10. New York, NY, USA: Springer, 2001.

[33] R. Moraes, J. F. Valiati, and W. P. G. Neto, ''Document-level sentiment classification: An empirical comparison between SVM and ANN,'' *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013.

[34] M. Pal and P. M. Mather, ''An assessment of the effectiveness of decision tree methods for land cover classification,'' *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, Aug. 2003.

[35] A. Glotov and P. Lyakhov, ''Pulmonary fibrosis progression prognosis using machine learning,'' in *Proc. Ural Symp. Biomed. Eng., Radioelectron. Inf. Technol. (USBEREIT)*, May 2021, pp. 327–329.

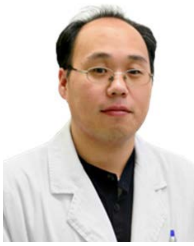
[36] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, ''An improved method to construct basic probability assignment based on the confusion matrix for classification problem,'' *Inf. Sci.*, vols. 340–341, pp. 250–261, May 2016.

[37] U. M. Fayyad and K. B. Irani, ''The attribute selection problem in decision tree generation,'' in *Proc. AAAI*, vol. 92, 1992, pp. 104–110.

[38] National Medical Center. *Statistics on Bed Utilization Rate, Ministry of Health and Welfare, South Korea*. Accessed: Nov. 20, 2021. [Online]. Available: https://mohw.go.kr

**EUNBI KIM** received the M.S. degree in business administration from Hanyang University, Seoul, South Korea, in 2011. She is currently pursuing the Ph.D. degree in industrial and management engineering with Korea University, Seoul.

Her research interests include hospital operations, simulation, machine learning, and stochastic optimization. She, nowadays, focuses on data-driven decision support for emergency department operations.



**KAP SU HAN** received the B.S. degree in medicine and the Ph.D. degree in emergency medicine from Korea University, Seoul, South Korea, in 2002 and 2015, respectively.

He is currently an Associate Professor with the College of Medicine, Korea University. His research interests include management of emergency medical service systems and critical care.

**TAESU CHEONG** received the B.S. degree in industrial engineering from Korea University, Seoul, South Korea, in 1998, the M.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2001, and the Ph.D. degree in industrial and systems engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011.

He is currently a Professor with the School of Industrial and Management Engineering, Korea University. His research interests include stochastic optimization with applications in transportation, supply chain management, healthcare management, and information system management.

**SUNG WOO LEE** received the B.S. degree in medicine, the M.S. degree in general surgery of medicine, and the Ph.D. degree in emergency medicine from Korea University, Seoul, South Korea, in 1993, 1999, and 2001, respectively.

He is currently a Professor with the School of Medicine, Korea University, and the Director of the Seoul Emergency Medical Support Center and the Seoul Poisoning Control Center. His research interests include clinical toxicology, and management of emergency medical service systems and critical care.

**JOONYUP EUN** received the B.S. degree in industrial systems and information engineering from Korea University, Seoul, South Korea, in 2007, the M.S. degree in industrial and systems engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2009, and the Ph.D. degree in industrial engineering from Purdue University, West Lafayette, IN, USA, in 2016.

He is currently an Assistant Professor with the Graduate School of Management of Technology, Korea University. Prior to joining Korea University, he was a Postdoctoral Research Fellow of operations research with the Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, USA. His research interests include data-driven modeling and analytics of hospital operations and capacity management based on deterministic/stochastic optimization of service systems.

**SU JIN KIM** received the B.S. degree in medicine and the M.S. and Ph.D. degrees in emergency medicine from Korea University, Seoul, South Korea, in 1999, 2003, and 2006, respectively.

She was a Visiting Fellow at the WISER Institute, University of Pittsburgh. She is currently a Professor of emergency medicine at Korea University. She is also conducting research projects on medical device development, AI-based analysis in emergency and critical care medicine. Her academic and research interests include resuscitation (ECPR), critical care medicine (hemodynamic and non-invasive monitoring), simulation-based training, and system quality improvement, including disaster preparedness.