

# Machine Learning and Initial Nursing Assessment-Based Triage System for Emergency Department

Jae Yong Yu<sup>1</sup>, Gab Yong Jeong<sup>2</sup>, Ok Soon Jeong<sup>3</sup>, Dong Kyung Chang<sup>1,3,4</sup>, Won Chul Cha<sup>1,2,3</sup>

<sup>1</sup>Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul, Korea

<sup>2</sup>Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

<sup>3</sup>Department of IT Planning, Samsung Medical Center, Seoul, Korea

<sup>4</sup>Department of Gastroenterology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

**Objectives:** The aim of this study was to develop machine learning (ML) and initial nursing assessment (INA)-based emergency department (ED) triage to predict adverse clinical outcome. **Methods:** The retrospective study included ED visits between January 2016 and December 2017 that resulted in either intensive care unit admission or emergency room death. We trained four classifiers using logistic regression and a deep learning model on INA and low dimensional (LD) INA, logistic regression on the Korea Triage and acuity scale (KTAS) and Sequential Related Organ Failure Assessment (SOFA). We varied the outcome ratio for external validation. Finally, variables of importance were identified using the random forest model's information gain. The four most influential variables were used for LD modeling for efficiency. **Results:** A total of 86,304 patient visits were included, with an overall outcome rate of 3.5%. The area under the curve (AUC) values for the KTAS model were 76.8 (74.9–78.6) with logistic regression and 74.0 (72.1–75.9) for the SOFA model, while the AUC values of the INA model were 87.2 (85.9–88.6) and 87.6 (86.3–88.9) with logistic regression and deep learning, suggesting that the ML and INA-based triage system result more accurately predicted the outcomes. The AUC values for the LD model were 81.2 (79.4–82.9) and 80.7 (78.9–82.5) for logistic regression and deep learning, respectively. **Conclusions:** We developed an ML and INA-based triage system for EDs. The novel system was able to predict clinical outcomes more accurately than existing triage systems, KTAS and SOFA.

**Keywords:** Triage, Machine Learning, Deep Learning, Hospital Emergency Service, Efficiency

**Submitted:** May 9, 2019

**Revised:** 1st, July 12, 2019; 2nd, August 14, 2019; 3rd, October 15, 2019

**Accepted:** October 17, 2019

## Corresponding Author

Won Chul Cha

Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81, Irwon-ro, Gangnam-gu, Seoul 06351, Korea. Tel: +82-2-3410-2053, E-mail: wc.cha@samsung.com (<https://orcid.org/0000-0002-2778-2992>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2020 The Korean Society of Medical Informatics

## 1. Introduction

An emergency department (ED) is a complex scene where various diseases and processes are intertwined. Annually, over 4.8 million patients visit EDs in Korea, and 137.8 million visit EDs in the United States [1,2]. Moreover, the number of patients and the severity of their complaints are increasing due to aging of the population and advances in emergency medicine [3]. When resources are not sufficient, the increased load on EDs results in a poor quality of care, which leads to a suboptimal outcome [4].

Triage systems have been developed where demand is greater than supply [5]. The purpose of triage in an ED is to prioritize patients to allocate clinical resources as beds

and providers. There are several triage systems worldwide: the Emergency Severity Index (ESI), the Korean Triage and Acuity Scale (KTAS), the Canadian Triage and Acuity Scale (CTAS), and so forth [6-8]. Studies have revealed that these systems are useful and necessary [9]. However, drawbacks such as human dependency and ambiguity of judgment have also been highlighted [10]. These potential problems could worsen when the volume of patients increases and information accumulates.

Digitalized triage systems have been introduced to support triage decisions by healthcare providers [10]. These systems have shown reliable outcomes in simulation settings, which are often compared to human decision. However, only limited value of such systems has been discovered [11].

One of the most important aspects of care in an ED is the initial assessment of a patient's condition. Nurses who encounter patients first measure their condition so that they can identify and manage their physical, mental, or social problems [12]. Regarding the fact that large amounts of data are gained at the moment of initial nursing assessment (INA), such as age, gender, initial vital signs, etc., using only the triage score for decision may be inefficient. Machine learning (ML) could provide an effective approach to utilize this information. ML is a method that allows a computer to train by itself from data without explicit coding. From a large amount of data, ML automatically learns the features or representations for a given task, such as classification, detection, or prediction [13].

The aim of this study was to evaluate an ML and INA-based ED triage system to predict adverse clinical outcomes.

## II. Methods

### 1. Study Setting

This study was a single-center, retrospective study, conducted in an ED of a tertiary academic hospital (a 1,960-bed, university-affiliated hospital located in a metropolitan city with an annual census of 70,000) [14].

### 2. Study Subject

The study subjects were defined as ED visitors from January 1, 2016 to December 31, 2017.

We excluded patients who were non-adult (age <18), were dead on arrival (DOA) or after cardiopulmonary resuscitation (CPR) or injury. Missing lab data were also excluded from Sequential Organ Failure Assessment (SOFA) score calculation. The process of selecting patients is illustrated in Figure 1.

The Institutional Review Board of Samsung Medical Center approved this study. Informed consent was exempted because this was a retrospective, observational, and de-identified study (No. SMC 2018-11-007).

### 3. Feature

Data were selected from a clinical data warehouse (CDW) detailing age, gender, level of consciousness, route of arrival, method of transportation, weekend, day of works, vital signs (temperature, heart rate, systolic blood pressure, respiratory rate, oxygen saturation), and initial KTAS score by nurse staff. There are two types of KTAS, namely, initial assessment and reassessment of the condition of a patient. We used the initial KTAS for this study. We also used the SOFA score with the partial pressure of oxygen ( $\text{PaO}_2$ ), fraction of inspired oxygen ( $\text{FiO}_2$ ) for respiration, platelet count for coagulation, bilirubin for liver, mean arterial pressure for the cardiovascular system, Glasgow Coma Scale (GCS) for the central nervous system, and creatinine or urine output for the renal system. The SOFA score calculates the number and severity of dysfunctions in six organ systems (respiration, coagulation, liver, cardiovascular, central nervous system, and renal). Each organ system is assigned a point value from 0 (normal) to 4 (high degree of dysfunction/failure) [15]. Using clinical experience, vital sign features were categorized into groups [10] and were used to build the model, such as

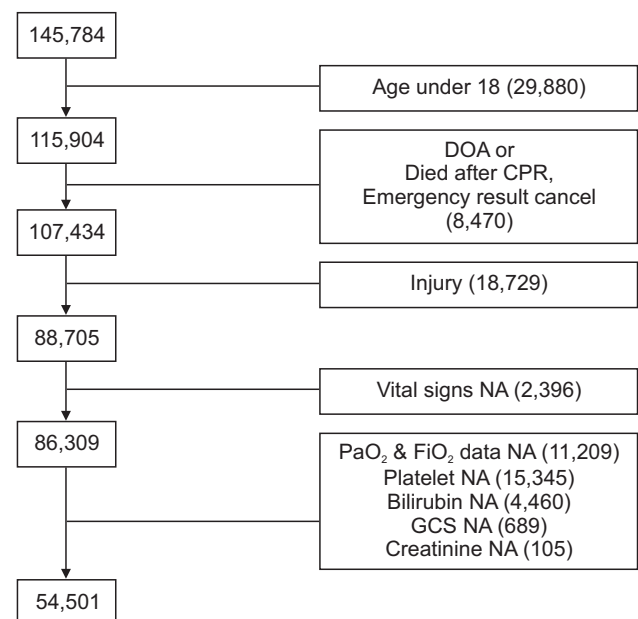


Figure 1. Study inclusion criteria for the emergency room cohort. DOA: dead on arrival, CPR: cardiopulmonary resuscitation, NA: not available (missing), GCS: Glasgow Coma Scale.

patient demographic information, vital signs, and emergency severity index, which is measured initially and is mandatory when patients visit an ED (Figure 2). Information regarding the ED is sent to the National Emergency Medical Center.

**4. Study Outcome**

Our primary and composite outcome was mortality in the ED or intensive care unit (ICU) admission. These clinical outcomes were included as a target feature for analysis to build the model.

**5. Model Development and Evaluation**

We built the prediction model to quantify the probability of clinical outcome. A patient’s likelihood of outcome may serve as a proxy for acuity which could be comparable with KTAS and SOFA scores.

All data processing and statistical analysis were conducted using R version 3.5.0 software (<https://www.r-project.org/>). We divided each feature into clinical classifications, with the cutoff values from previous research [10].

We divided the patients into three sets, namely, training, validation, and test sets, for modeling, model parameter tuning, and evaluation, respectively.

Multivariate logistic regression analysis was conducted using R package ‘glm’ to estimate the likelihood of clinical outcomes after adjusting for outcome ratio and other potential factors that can determine which variables had the greatest

effect on outcome. Further, an ML method known for good classification was used, namely, deep learning with R package ‘Keras’. The following hyper-parameters were used: number of layers and number of hidden units in deep learning, which were validated and selected using the validation set. Receiver operating characteristic (ROC) curves were generated by varying the thresholds of each model prediction probability. Finally, several models were compared, and the best prediction models were selected based on their area under the ROC curve (AUROC) values. The AUROC value for the model and its confidence interval were expressed with a 95% confidence interval (CI). We also used variable importance plots in random forest to determine which variables affect the results. Chief complaint, age, heart rate, and SpO<sub>2</sub> were the most influential factors for predicting clinical outcomes. Those variables were used for low dimensional (LD) modeling for efficiency.

To compare the acuity of the patients, we cut the model likelihood for the individual patients with the KTAS level ratio to make a model-based KTAS with the same ratio. We compared KTAS with ML-based KTAS. In addition to the contingency table comparison, we show a matrix heatmap for comparison of two KTAS.

Because there was a class imbalance in which less than 15% of the total cases were positive, we considered the Synthetic Minority Over-sampling Technique (SMOTE) to solve this imbalance problem [16]. We sampled minority and majority class

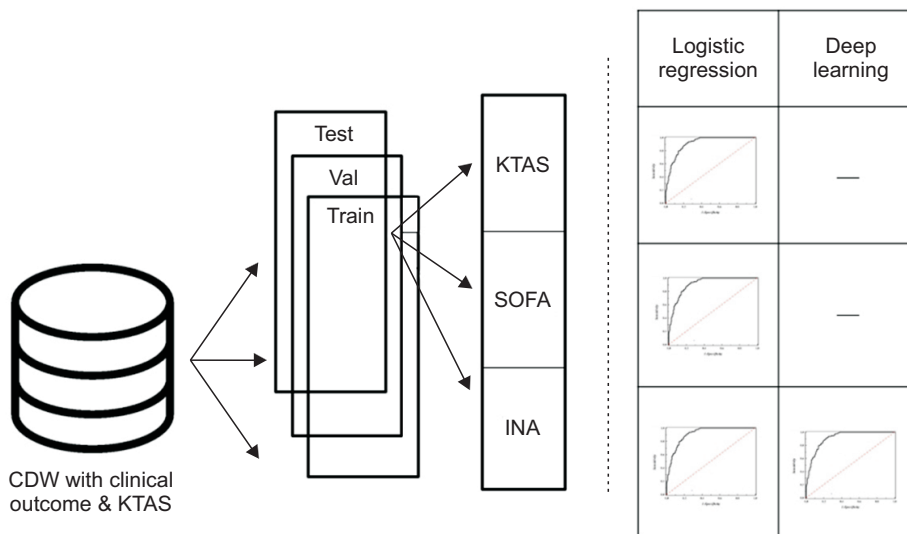


Figure 2. Overview of workflow. Clinical data warehouse (CDW) with clinical outcome is split into three parts (training, validation, and testing). Three different methods (logistic regression, random forest, and deep learning) and two different types of input (KTAS and INA without KTAS) were combined. Model evaluation was judged by AUROC. KTAS: Korea Triage and Acuity Scale, SOFA: Sequential Organ Failure Assessment, INA: initial nursing assessment, AUROC: area under the receiver operating characteristic curve.

Table 1. Demographic and clinical characteristics of ER patients

	Train (n = 51,785)	Validation (n = 17,262)	Test (n = 17,262)
Age (yr)	55.6 ± 17.7	55.9 ± 17.6	55.6 ± 17.6
Gender			
Male	25,240 (48.7)	8,503 (49.3)	8,486 (49.2)
Female	26,545 (51.3)	8,759 (50.7)	8,776 (50.8)
Level of consciousness			
Alert	50,419 (97.4)	16,803 (97.3)	16,831 (97.5)
Non-alert <sup>a</sup>	1,366 (2.6)	459 (2.7)	431 (2.5)
Route of arrival			
Direct visit	41,211 (79.6)	13,817 (80.0)	13,802 (80.0)
Referred from outside	7,121 (13.8)	2,312 (13.4)	2,368 (13.7)
Referred from OPD	3,239 (6.3)	1,074 (6.2)	1,033 (6.0)
Others	214 (0.4)	59 (0.3)	59 (0.3)
Method of transportation			
Ambulance	10,589 (20.4)	3,598 (20.8)	3,513 (20.4)
Others	41,196 (79.6)	13,664 (79.2)	13,749 (79.6)
Day of week			
Weekend	14,682 (28.4)	4,934 (28.6)	4,997 (28.9)
Non-weekend	37,103 (71.6)	12,328 (71.4)	12,265 (71.1)
Hours			
Office hour (9–18)	27,323 (52.8)	9,109 (52.8)	8,960 (51.9)
Non-office hour	24,462 (47.2)	8,153 (47.2)	8,302 (48.1)
Vital signs			
Body temperature (°C)			
Normal (35.0–38.0)	46,116 (89.1)	15,381 (89.1)	15,450 (89.5)
Abnormal	5,669 (10.9)	1,881 (10.9)	1,812 (10.5)
Heart rate (bpm)			
Normal (50.0–110.0)	43,156 (83.3)	14,350 (83.1)	14,429 (83.6)
Abnormal	8,629 (16.7)	2,912 (16.9)	2,833 (16.4)
SBP (mmHg)			
Normal (100.0–200.0)	46,422 (89.6)	15,484 (89.7)	15,479 (89.7)
Abnormal	5,363 (10.4)	1,778 (10.3)	1,783 (10.3)
Respiratory rate (breaths/min)			
Normal (14.0–19.0)	27,592 (53.3)	9,220 (53.4)	9,172 (53.1)
Abnormal	24,193 (46.7)	8,042 (46.6)	8,090 (46.9)
Oxygen saturation (%)			
Normal (95.0–100.0)	48,178 (93.0)	16,082 (93.2)	16,041 (92.9)
Abnormal	3,607 (7.0)	1,180 (6.8)	1,221 (7.1)

Values are presented as mean ± standard deviation or number (%).

ER: emergency room, OPD: outpatient department, SBP: systolic blood pressure.

<sup>a</sup>Non-alert: verbal, painful, unconscious.

values from a 1:1 ratio to a 1:3 ratio using R package ‘DMwR’.

Descriptive statistics were used for the demographic features and characteristics of the ED visits. Categorical variables are expressed in counts and percentages of the total amount of data available within the database.

### III. Results

The initial data included 145,784 ED visits by individuals aged  $\geq 18$  ( $n = 115,904$ ) and excluded those who were DOA or had died after CPR, cancelled cases ( $n = 107,434$ ), and injury ( $n = 88,705$ ). The data were filtered by excluding cases that had missing vital sign information ( $n = 2,396$ ). Finally, data on 86,309 ED visits were included in the study. There were 51,785 (60.0%) patients in the training set, 17,262 (20.0%) in the validation set, and 17,262 (20.0%) in the testing set.

The distribution of ED patients demographics divided into three groups is shown in Table 1. Of the 86,309 patients, 157 (0.18%) died, 3,024 (3.50%) were transferred to the ICU or died during their ED stay. The number of female patients (51.1%) was greater than the number of male patients visiting the ED from 2016 to 2017, but there was no statistically significant difference in proportion,  $\chi^2$  ( $df = 2$ ) = 1.86,  $p = 0.395$ . With regard to the level of consciousness, 97.4% of the patients were alert at the time of the ED visit. A total of 79.7% of ED visits were direct visits, and 13.7% were referred. Approximately a quarter (20.5%) of the patients used an ambulance. The proportion of KTAS level 3 was highest (46.8%), the other proportion cases with KTAS levels of 1, 2, 4, and 5 were 0.60%, 9.07%, 35.90%, and 7.66%, respectively. With regard to the Vital signs, most patients were normal. The range of normal proportion was from 53.3% to 92.9%.

The outcomes among different severity groups were analyzed and compared separately. Table 2 shows the distribution of ED patients among clinical outcome and KTAS. The proportions differed between severity groups. In the admission to ICU group, levels 1 and 2 are the most common. In the admission to ward group, levels 2 and 3 are the most common. The proportion of death and transfer decreased in the KTAS level order. In contrast, the proportion of discharge increased in the KTAS level order. We also analyzed the frequency of chief complaint as well as the frequency of the major chief complaint symptoms, mostly abdominal pain (22.1%), fever (16.5%), and dyspnea (11.8%), as seen in Supplementary Table S1.

The AUROC values for KTAS and SOFA only model were 76.8 (74.9–78.6) and 74.0 (72.1–75.9) with logistic regression, while the AUROC values for the INA model were 87.2 (85.9–88.6) and 87.6 (86.3–88.9) with logistic regression and deep learning, respectively, suggesting that the INA-based triage system result more accurately predicted outcomes. The AUC values for the LD model using the most four influential features were 81.2 (79.4–82.9) and 80.7 (78.9–82.5) with logistic regression and deep learning, respectively, indicating that the efficiency model also outperformed the KTAS and SOFA models (Table 3).

Varying the outcome ratio from 1 to 3 showed consistent results as shown in Supplementary Table S2.

Figure 3 shows the difference in the distribution of triage results of the KTAS and ML-based triage scores. The cutoff value of the ML-based triage was determined according to the proportion of triage scores of the KTAS. The results show the inter-reliability of the ML-based model compared to KTAS with estimated Cohen’s kappa statistics  $k = 0.0018$ ,

Table 2. ED disposition & outcome for KTAS level

	KTAS				
	Level 1	Level 2	Level 3	Level 4	Level 5
<b>ED disposition</b>					
Admission ICU	159 (30.69)	1,113 (14.21)	1,280 (3.17)	285 (0.92)	30 (0.45)
Admission ward	192 (37.07)	2,494 (31.85)	12,765 (31.62)	6,661 (21.50)	959 (14.50)
Death	27 (5.21)	50 (0.64)	64 (0.16)	15 (0.05)	1 (0.02)
Discharge	91 (17.57)	3,910 (49.93)	25,161 (62.34)	23,450 (75.68)	5,554 (84.00)
Transfer	49 (9.46)	264 (3.37)	1,094 (2.71)	573 (1.85)	68 (1.03)
<b>Outcome</b>					
Mortality or ICU admission	186 (35.91)	1,163 (14.85)	1,344 (3.33)	300 (0.97)	31 (0.47)
Overall	518	7,831	40,364	30,984	6,612

Values are presented as number (%).

ED: emergency department, KTAS: Korea Triage and Acuity Scale, ICU: intensive care unit.

**Table 3. Summary of AUROC (95% confidence interval) of each method**

	Logistic regression	Deep learning
KTAS	76.8 (74.9–78.6)	-
SOFA	74.0 (72.1–75.9)	-
INA	87.2 (85.9–88.6)	87.6 (86.3–88.9)
LD INA	81.2 (79.4–82.9)	80.7 (78.9–82.5)

AUROC: area under the receiver operating characteristic curve, KTAS: Korea Triage and Acute Scale, SOFA: Sequential Organ Failure Assessment, INA: initial nursing assessment, LD INA: low-dimensional initial nursing assessment.

suggesting that the extent of agreement is slight.

Variable importance plots in random forest were also used to estimate the impact of features. The results are shown in Supplementary Figure S1. Chief complaint, age, HR, and SpO<sub>2</sub> were the most important factors in the model.

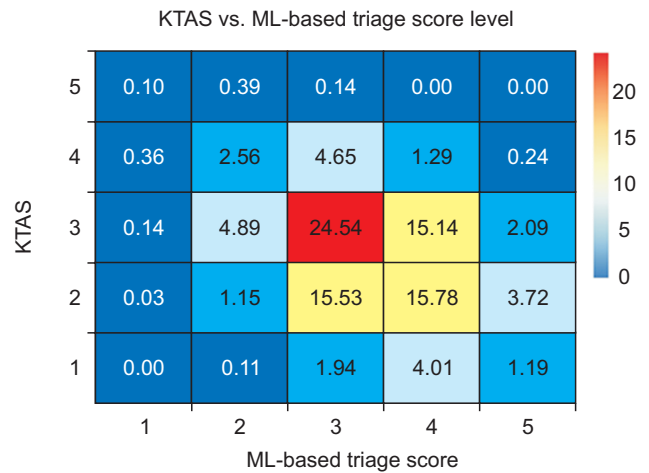
### IV. Discussion

There were some limitations of this study. First, an external validation was not performed. Patients’ characteristics differ among institutions, and additional learning would be needed to generalize the algorithm from this study. To test the performance of the algorithm with a population of varying severity, we used a SMOTE method for over-sampling and under-sampling.

Second, as it is the result of the initial characteristics of the patient visiting the emergency room (ER), it did not reflect the physicians’ notes or lab results generated later. However, our study aimed to determine which process the triage system would put the patient into, so the analysis of the information generated later will be done in the next study.

Finally, a cross-sectional study does not reflect medical history. We can set the time window and consider the medical history during the time interval. For example, we can consider the number of hospitalizations, outpatient visits, procedures or surgeries during 2 years. It is necessary to study the model that predicts the outcome by further utilizing the patient’s medical history information.

In this study, we developed an ML tool and evaluated its performance in EDs, using KTAS and SOFA as a comparator. Our results show that INA-ML is the most suitable for clinical outcome (ER death or ICU admission) prediction. However, it is in the model type-free result that we can see there was no significant difference from logistic regression or other ML methods.



**Figure 3. Ratio for nursing Korea Triage and Acuity Scale (KTAS) and machine learning (ML)-based KTAS.**

Once we can estimate the effect of each feature in the model, it can be easily incorporated into electronic medical record system and can immediately calculate a score once common patient information has been put into the system. Our model was designed for clinical decision support, and it is calculated immediately based on the first data recorded on a patient’s arrival. It is not intended to totally replace nurse or physician judgement, but to support them by providing the score as they assess a patient.

INA can be used for two purposes. KTAS can be used for measuring the patient proxy severity and INA for common patient information. Additional information can be obtained when we combined these two attributes. The advantages of the combined model over the reference standard KTAS include evidence-based development and decreased reliance on subjective human experience or judgement [13].

Even though speed and accuracy are the most important factors in ER results, not many studies on the initial information of emergency patients have been conducted. Second, we did not use additional information because there was no tool for this. We can upgrade the triage by making good use of structured data through this study. Third, we did not use only one method [11] and compared various methodologies. Logistic regression alone has limitations, so we compared random forest and deep learning. Random forest and deep learning made good predictions. Especially when we checked the contents of the variables, the main symptom and vital sign had a great influence on the prediction. Despite these strengths of the ML method, we need to decide which parameter is the best. To solve this parameter setting problem, we split the dataset into training, validation, and testing sets. The ranges of parameters were selected by human

decision. If we broaden the range of parameter or use other optimization methods, such as random search, grid search, and so forth, we could get more predictable parameters and outcomes. Finally, study should be made for practical applications. The reason for making the model is for efficiency in the actual clinical field, like Watson for Oncology. Therefore, it is necessary to integrate the model and data for practical use. Many variables can produce good prediction results, but it could be problematic to apply them in the medical environment. It could be a proxy way to show the results using some LD variables with high variable importance in clinical settings.

We have developed an ML and INA-based triage system for EDs. The novel system was able to predict clinical outcomes more accurately than existing triage systems, namely, KTAS and SOFA.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## ORCID

Jae Yong Yu (<https://orcid.org/0000-0001-6922-8538>)

Gab Yong Jeong (<https://orcid.org/0000-0002-0998-6828>)

Ok Soon Jeong (<https://orcid.org/0000-0002-6172-6329>)

Dong Kyung Chang (<https://orcid.org/0000-0001-8925-4629>)

Won Chul Cha (<https://orcid.org/0000-0002-2778-2992>)

## Supplementary Materials

Supplementary materials can be found via <http://doi.org/10.4258/hir.2020.26.1.13>. Table S1. Chief complaint frequency table. Table S2. Varying clinical outcome ratio and AUROC. Figure S1. Variable importance plot from random forest.

## References

- Kang SW, Park HS. Emergency department visit volume variability. *Clin Exp Emerg Med* 2015;2(3):150-4.
- Moore BJ, Stocks C, Owens PL. Trends in emergency department visits, 2006–2014 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality; 2017 [cited 2020 Jan 10]. Available from: <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb227-Emergency-Department-Visit-Trends.pdf>.
- Kasteridis P, Street A, Dolman M, Gallier L, Hudson K, Martin J, et al. Who would most benefit from improved integrated care? Implementing an analytical strategy in South Somerset. *Int J Integr Care* 2015;15:e001.
- Combes JR, Arespachoga E. Appropriate use of medical resources [Internet]. Chicago (IL): American Hospital Association's Physician Leadership Forum; 2013 [cited at 2020 Jan 10]. Available from: <http://www.aha-physicianforum.org/files/pdf/appropusewhiteppr.pdf>.
- Robertson-Steel I. Evolution of triage systems. *Emerg Med J* 2006;23(2):154-5.
- Schull MJ, Kiss A, Szalai JP. The effect of low-complexity patients on emergency department waiting times. *Ann Emerg Med* 2007;49(3):257-264.e1.
- Silva JA, Emi AS, Leao ER, Lopes MC, Okuno MF, Batista RE. Emergency Severity Index: accuracy in risk classification. *Einstein (Sao Paulo)* 2017;15(4):421-7.
- Ajani K. Triage: a literature review of key concepts. *J Pak Med Assoc* 2012;62(5):487-9.
- FitzGerald G, Jelinek GA, Scott D, Gerdtz MF. Emergency department triage revisited. *Emerg Med J* 2010;27(2):86-92.
- Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. *J Emerg Med* 2016;50(6):910-8.
- Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. *Ann Emerg Med* 2018;71(5):565-574.e2.
- Edmonds J. Initial assessment of young people in the emergency department. *Emerg Nurse* 2011;19(8):32-6.
- Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J. Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS One* 2018;13(10):e0205836.
- Hwang SY, Jo IJ, Lee SU, Lee TR, Yoon H, Cha WC, Sim MS, Shin TG. Low accuracy of positive qSOFA criteria for predicting 28-day mortality in critically ill septic patients during the early period after emergency department presentation. *Ann Emerg Med* 2018;71(1):1-9.e2.
- Baig MA, Sheikh S, Hussain E, Bakhtawar S, Subhan Khan M, Mujtaba S, et al. Comparison of qSOFA and SOFA score for predicting mortality in severe sepsis and septic shock patients in the emergency department of a low middle income country. *Turk J Emerg Med* 2018;18(4):148-51.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-57.