

Machine Learning–Based Prediction of Korean Triage and Acuity Scale Level in Emergency Department Patients

Sae Won Choi^{1,2}, Taehoon Ko¹, Ki Jeong Hong^{2,3}, Kyung Hwan Kim^{1,4}

¹Office of Hospital Information, Seoul National University Hospital, Seoul, Korea

²Department of Emergency Medicine, Seoul National University Hospital, Seoul, Korea

³Laboratory of Emergency Medical Services, Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

⁴Department of Thoracic and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul, Korea

Objectives: Triage is a process to accurately assess and classify symptoms to identify and provide rapid treatment to patients. The Korean Triage and Acuity Scale (KTAS) is used as a triage instrument in all emergency centers. The aim of this study was to train and compare machine learning models to predict KTAS levels. **Methods:** This was a cross-sectional study using data from a single emergency department of a tertiary university hospital. Information collected during triage was used in the analysis. Logistic regression, random forest, and XGBoost were used to predict the KTAS level. **Results:** The models with the highest area under the receiver operating characteristic curve (AUROC) were the random forest and XGBoost models trained on the entire dataset (AUROC = 0.922, 95% confidence interval 0.917–0.925 and AUROC = 0.922, 95% confidence interval 0.918–0.925, respectively). The AUROC of the models trained on the clinical data was higher than that of models trained on text data only, but the models trained on all variables had the highest AUROC among similar machine learning models. **Conclusions:** Machine learning can robustly predict the KTAS level at triage, which may have many possibilities for use, and the addition of text data improves the predictive performance compared to that achieved by using structured data alone.

Keywords: Triage, Hospital Emergency Service, Machine Learning, Natural Language Processing

Submitted: October 8, 2019

Revised: October 21, 2019

Accepted: October 21, 2019

Corresponding Author

Kyung Hwan Kim

Office of Hospital Information, Seoul National University Hospital,
101 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Tel: +82-2-2072-
7600, E-mail: kkh726@snu.ac.kr ([https://orcid.org/0000-0002-
2718-8758](https://orcid.org/0000-0002-2718-8758))

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.

© 2019 The Korean Society of Medical Informatics

I. Introduction

Ten million patients visited about 400 emergency departments in 2018 in Korea, and for most of those patients, the first emergency care they encountered was triage [1]. Triage, implemented in most modern emergency departments, is the process of accurately assessing and classifying severe symptoms to identify and provide rapid treatment to emergency patients. To this end, a patient's vital signs are measured, and a short history and examination is carried out as soon as the patient arrives at the emergency department by emergency physicians, emergency specialist nurses, or emergency medical technicians.

The Korean Triage and Acuity Scale (KTAS) was developed in 2012 based on the Canadian Triage and Acuity Scale [2].

Since 2016, it has been mandatory for all emergency centers across the country to use and assess patients visiting emergency departments with the KTAS. Centers are incentivized by the National Health Insurance Service through higher reimbursement rates for patients with higher KTAS levels [3]. Depending on the assigned KTAS level, a patient may wait for a set maximum amount of time until being evaluated by a physician [2]. It also supports the prioritization of the care provided in a busy emergency department to patients requiring urgent care. The KTAS was also recently shown to improve the length of stay and mortality rates in emergency departments [4].

As with most triage instruments, the KTAS classifies patients into five levels based on the patients' symptoms, with primary and secondary factors taken into account [5]. Primary factors involve characteristics common to most symptoms, such as consciousness, blood pressure, heart rate, respiration rate, fever, pain, presence of hemorrhage, and trauma. Secondary considerations are characteristics applied to specific symptoms. Although most criteria are objective indicators, subjective judgment by the triage staff influences the final level assigned to the patient.

Since triage is the starting point of care for emergency patients, it is important to be consistent in assigning KTAS levels. Hospitals employ dedicated personnel who have been certified by the KTAS Committee under the Korean Society of Emergency Medicine. The committee maintains a 6-hour training program for providers with more than 1-year clinical experience in an emergency department [6]. It is also important to maintain the quality of education, training, and evaluation, which is becoming more difficult to maintain, as the complexity of emergency care increases and more patients visit emergency departments nationwide. However, even with these systems in place, the problems of misclassification, over-triage, and under-triage still remain, due to the inherent complexity and uncertainty of triage.

Several previous studies have applied machine learning models to patients in intensive care units [7,8], and emergency departments [9,10]. Many studies have considered triage in particular, with most aiming to predict the outcome at triage [11–14]. One such example is a study that compared e-triage with the emergency severity index that showed more accurate results in predicting patient outcome [15]. Other efforts have focused on using not just structured data but also the text data generated. Sterling et al. [16] predicted disposition from the emergency department with only the triage text, and Hong et al. [11] used information collected at triage in addition to textual patient history to predict admission

from the emergency department. However, there have been no attempts to predict or determine the outcome of a triage instrument with information collected during triage.

The aim of this study was to train and compare machine learning models in their ability to predict KTAS levels. We also hypothesized that free text nursing notes written during triage would provide additional information that would support prediction of the KTAS level.

II. Methods

1. Study Design, Setting and Data Source

This study was a cross-sectional study using data from a single emergency department of a tertiary university hospital with an annual census of 60,000 patients. Retrospective data of all visits by patients to the adult emergency department from November 2016 to June 2019 were included in the study. Encounters with missing data were excluded. All data processing was done with Python and related scientific libraries, pandas, and scikit-learn [17,18]. This study and its protocol were approved by the Institutional Review Board of Seoul National University Hospital (No. 1910-071-1070) with a waiver of informed consent. No personally identifiable data was part of the dataset. Data used in this study was retrieved from the clinical data warehouse of Seoul National University Hospital Patients Research Environment (SU-PREME).

2. Variables

Only variables collected during triage were used in the analysis. The data were divided into three separate datasets: structured data only, text of nursing triage notes only, and both structured data and nursing triage notes. The structured data included gender, age, date and time of arrival, chief complaint, route to the emergency department, pain location and intensity, vital signs (blood pressure, heart rate, respiratory rate, oxygen saturation), and level of consciousness. The chief complaint was selected from a set of 607 codes; the route taken by the patient to the emergency department was selected from five options (direct, refer from outpatient department, transfer from other hospital, other, or unknown); the pain location was selected from 10 possible locations; level of consciousness was selected from five options (alert, verbal response-disoriented, verbal response-stuporous, response to pain, and unresponsive). Some additional preprocessing was applied to the data. Arrival time was binned into 1-hour increments. Pain intensity scores that were written as text were converted into numeric form,

for instance '6-7' was converted to 6.5. The nursing triage note is a free-form text note, usually one to three sentences in length that summarizes the patient's reason for visiting the emergency department written by the triage nurse. The outcome variable was the KTAS level assigned to the patient by the triage nurse on duty.

3. Natural Language Processing

Natural language processing (NLP) is a set of methods for analyzing human languages. Its purpose is to allow computers to process and understand the meaning of text. Because each language has specialized structures, such as characters, grammar, and words, NLP techniques have been developed for each specific language. In this study, most of the targeted documents were written in Korean. The domain of documents should also be considered when implementing NLP. The targeted documents were triage notes containing both Korean and English words. Most of them followed Korean grammar but had many English words designating medical concepts, such as symptoms, medications, and disease names.

To analyze the triage notes, we used `soynlp`, which is an NLP library of unsupervised NLP techniques written and available in Python [19]. `Soynlp` was used to tokenize the triage notes and normalize words. Unlike Western languages like English, Korean is not divided into meaning structures by spaces alone. For this reason, most Korean morpheme analyzer tools are based on pre-defined dictionaries. However, it is very difficult for these tools to analyze text that contain words that are not in the dictionary. This is called an out-of-vocabulary problem. In contrast, `soynlp` uses a word extractor based on probabilistic scores representing the likelihood of words. Therefore, it is more appropriate for analyzing documents with terminology that is used only in specific domains compared to any other Korean processing tools.

Because of the time pressure involved in writing the initial triage note, they usually contain abbreviations and shorthand notations. Because punctuation marks and numbers are usually removed in the preprocessing step before analysis, to maintain the meaning of notations that include multiple punctuation marks, the most common shorthand notations were fully spelled out. For example, abbreviations such as 'n/v' were replaced with 'nausea vomiting' and notations such as '(+/-)', which would be completely removed by the preprocessing step removing punctuation, was replaced with the words 'positive negative'. There was a total of 165 such replacement rules. After these abbreviations and shorthand notations had been removed, punctuation and numbers were

removed, and all English text was converted into lowercase. No special preprocessing of the Korean language was performed besides tokenization with `soynlp`.

4. Machine Learning Models

In this study, we used logistic regression, random forest, and XGBoost to classify encounters. We applied a ridge logistic regression which uses L2 penalty to avoid multicollinearity and overfitting. Random forest is an ensemble model to combine predicted results of multiple decision trees [20]. In the process of learning a random forest model, each decision tree trains on a random sample of the training dataset with replacement, and a random subset of features is selected at each candidate split of trees. Through these methods, the trained trees become mostly uncorrelated. Averaging the results of less correlated models has the effect of reducing the high variance of each model's performance. As a result, the random forest has improved generalization performance. Random forest has been used by many scientists due to its performance; it was the best model for even the most comprehensive benchmarking studies [21]. XGBoost is an advanced model of gradient boosting machines (GBM), which trains many decision trees in a sequential and additive manner [22]. Although both GBM and random forest are based on decision trees, they train trees in different ways. In random forest, each tree is trained independently and becomes a so-called fully-grown tree, which overfits a subset of training dataset. GBM, on the other hand, learns small trees which may underfit the training set. However, trees in GBM are trained sequentially, and each tree is trained in consideration of the errors of other tree models in the previous steps. In this way, as more trees are trained, the error of GBM is reduced. XGBoost makes GBM's training scalable by distributed learning with several approximation methods to parallelize training. Since XGBoost was proposed, it has outperformed other algorithms in many studies and competitions [23].

To apply these algorithms, we structured the clinical notes using a bag-of-words model. In a bag-of-words model, each document is represented as a vector that has the same length as the size of the vocabulary [24]. In this study, the vocabulary was composed of unigrams, bigrams, and trigrams appearing in more than 0.1% in the entire document set [25]. However, in preliminary experiments, the use of unigrams alone yielded better performance. Using bigrams and trigrams caused the document vector represented by the bag-of-words model to become too long, leading the machine learning algorithms to fall victim to the curse of dimension-

Table 1. Patient demographics

	KTAS level					Total
	1	2	3	4	5	
Number of patients	1,989 (1.44)	16,098 (11.66)	77,720 (56.31)	36,045 (26.12)	6,170 (4.47)	138,022
Gender						
Female	814 (40.9)	7,328 (45.5)	40,612 (52.3)	19,703 (54.7)	3,652 (59.2)	72,109 (52.2)
Male	1,175 (59.1)	8,770 (54.5)	37,108 (47.7)	16,342 (45.3)	2,518 (40.8)	65,913 (47.8)
Age (yr)						
0–29	76 (3.8)	1,056 (6.6)	7,477 (9.6)	5,961 (16.5)	1,433 (23.2)	23,563 (17.1)
30–39	76 (3.8)	1,050 (6.5)	7,384 (9.5)	4,526 (12.6)	819 (13.3)	12,690 (9.2)
40–49	143 (7.2)	1,627 (10.1)	8,470 (10.9)	4,378 (12.1)	692 (11.2)	27,390 (19.8)
50–59	280 (14.1)	2,678 (16.6)	13,667 (17.6)	5,993 (16.6)	945 (15.3)	16,003 (11.6)
60–69	425 (21.4)	3,769 (23.4)	17,197 (22.1)	6,752 (18.7)	1,068 (17.3)	13,855 (10.0)
70–79	586 (29.5)	3,918 (24.3)	16,247 (20.9)	5,784 (16.0)	855 (13.9)	29,211 (21.2)
≥80	403 (20.3)	2,000 (12.4)	7,278 (9.4)	2,651 (7.4)	358 (5.8)	15,310 (11.1)
Time of visit to ED						
09:00–18:00	1,058 (53.2)	9,259 (57.5)	43,755 (56.3)	18,511 (51.4)	3,216 (52.1)	75,799 (54.9)
18:00–09:00	931 (46.8)	6,839 (42.5)	33,965 (43.7)	17,534 (48.6)	2,954 (47.9)	62,223 (45.1)
Route to ED						
Direct visit	1,377 (69.2)	10,834 (67.3)	60,005 (77.2)	31,143 (86.4)	5,431 (88.0)	108,790 (78.8)
Refer from OPD	113 (5.7)	1,434 (8.9)	6,035 (7.8)	1,192 (3.3)	210 (3.4)	8,984 (6.5)
Transfer from other hospital	489 (24.6)	3,820 (23.7)	11,637 (15.0)	3,700 (10.3)	528 (8.6)	20,174 (14.6)
Others	10 (0.5)	7 (0.0)	32 (0.0)	5 (0.0)	0 (0.0)	54 (0.0)
Unknown	0 (0.0)	3 (0.0)	11 (0.0)	5 (0.0)	1 (0.0)	20 (0.0)
Pain location						
Head	8 (0.4)	575 (3.6)	6,338 (8.2)	1,420 (3.9)	106 (1.7)	8,447 (6.1)
Neck	2 (0.1)	55 (0.3)	800 (1.0)	341 (0.9)	36 (0.6)	1,234 (0.9)
Thorax	52 (2.6)	1,809 (11.2)	3,452 (4.4)	638 (1.8)	41 (0.7)	5,992 (4.3)
Abdomen	84 (4.2)	930 (5.8)	14,760 (19.0)	2,550 (7.1)	69 (1.1)	18,393 (13.3)
Back	2 (0.1)	39 (0.2)	755 (1.0)	242 (0.7)	25 (0.4)	1,063 (0.8)
Pelvis	8 (0.4)	80 (0.5)	2,533 (3.3)	693 (1.9)	46 (0.7)	3,360 (2.4)
All extremity	1 (0.1)	18 (0.1)	213 (0.3)	283 (0.8)	31 (0.5)	546 (0.4)
Upper extremity	5 (0.3)	67 (0.4)	1,086 (1.4)	2,196 (6.1)	204 (3.3)	3,558 (2.6)
Lower extremity	18 (0.9)	125 (0.8)	2,008 (2.6)	4,280 (11.9)	326 (5.3)	6,757 (4.9)
Others	16 (0.8)	244 (1.5)	2,085 (2.7)	2,401 (6.7)	301 (4.9)	5,047 (3.7)
None	1,793 (90.1)	12,156 (75.5)	43,690 (56.2)	21,001 (58.3)	4,985 (80.8)	83,625 (60.6)
Level of consciousness						
Alert	1,115 (56.1)	14,439 (89.7)	75,950 (97.7)	35,644 (98.9)	6,128 (99.3)	133,276 (96.6)
Verbal response (disoriented)	163 (8.2)	910 (5.7)	1,423 (1.8)	287 (0.8)	26 (0.4)	2,809 (2.0)
Verbal response (stuporous)	113 (5.7)	472 (2.9)	253 (0.3)	74 (0.2)	6 (0.1)	918 (0.7)
Pain response	196 (9.9)	266 (1.7)	85 (0.1)	38 (0.1)	4 (0.1)	589 (0.4)
Unresponsive	402 (20.2)	11 (0.1)	9 (0.0)	2 (0.0)	6 (0.1)	430 (0.3)

Values are presented as number (%) or median (interquartile range).

KTAS: Korean Triage and Acuity Scale, ED: emergency department, OPD: outpatient department.

ality.

Each of the models was trained and evaluated on three separate datasets: structured data only, text of the nursing triage notes only, and both structured data and nursing triage notes. Precision, recall, F1-score, and area under the receiver operating characteristic (AUROC) curve with 95% confidence interval (CI) are reported.

III. Results

A total of 142,972 patients visited the adult emergency department during the study period. Encounters with missing information (4,950 cases, 3.5%) were excluded from the study. The most common missing information was the location of pain (4,404 cases) followed by oxygen saturation (215 cases). After exclusion of the cases with missing information, a total of 138,022 patients were included in the study. The basic demographics of the patients can be seen in Table 1. The median age of the patients was 60 years (interquartile range, 43–72) and 52.2% were female. The KTAS levels were distributed as follows: 1,989 level 1 (1.4%), 16,098 level 2 (11.7%), 77,720 level 3 (56.3%), 36,045 level 4 (26.1%), and 6,170 level 5 (4.5%).

Table 2 and Figure 1 present the classification results of the machine learning models on different subsets of data. The models with the highest AUROC were the random forest and XGBoost models trained on the entire dataset (AUROC = 0.922, 95% CI 0.917–0.925 and AUROC = 0.922, 95% CI 0.918–0.925, respectively). For each model, a similar trend is noticeable. The AUROC of the models trained on the clinical data is higher than that of the models trained on the text data only, but the models trained on all variables showed the highest AUROC among similar models.

IV. Discussion

This study developed and compared several machine learning models using information collected during triage to predict the KTAS level. Our results indicate that machine learning algorithms can predict KTAS levels robustly during emergency department triage. To our knowledge, this is the first study to predict KTAS levels using information available during triage. Most of the studies reported in the literature have focused on predicting patient outcomes, but there are several possible uses of predicting not only the outcome but also the triage level.

As with many recent results, the random forest and XG-

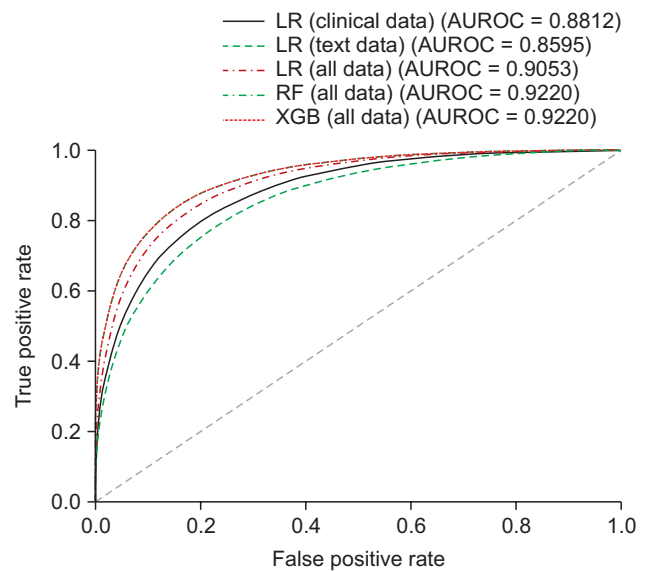


Figure 1. Receiver operating characteristic curve for selected machine learning models. AUROC: area under the receiver operating characteristic curve, LR: logistic regression, RF: random forest, XGB: XGBoost.

Table 2. Classification results of machine learning models

Model	Input variable	Precision	Recall	F1-score	AUROC (95% CI)
Logistic regression	Clinical data	0.680	0.659	0.666	0.881 (0.874–0.887)
	Text data	0.664	0.644	0.652	0.860 (0.851–0.868)
	All data	0.711	0.702	0.706	0.905 (0.900–0.909)
Random forest	Clinical data	0.743	0.744	0.729	0.913 (0.906–0.918)
	Text data	0.675	0.669	0.631	0.844 (0.833–0.853)
	All data	0.737	0.730	0.707	0.922 (0.917–0.925)
XGBoost	Clinical data	0.744	0.744	0.728	0.912 (0.906–0.917)
	Text data	0.664	0.661	0.623	0.842 (0.832–0.850)
	All data	0.753	0.753	0.740	0.922 (0.918–0.925)

AUROC: area under the receiver operating characteristic curve, CI: confidence interval.

Boost models outperform other machine learning algorithms, with XGBoost having the best F1-score (=0.740). It is also interesting to note that the AUROC of the random forest and XGBoost models using only clinical data outperformed the logistic regression model using the entire dataset pointing to the complex nonlinear relationships underlying the structured data.

As we hypothesized, the addition of nursing triage text data improved the prediction performance of all of the models studied, suggesting that there is information in the textual data that is not captured by the structured triage data, even with the minimal preprocessing of the text performed in this study. We only spelled out some common abbreviations and shorthand notations used in the notes, removed punctuation and transformed all English words to lowercase. We did not perform stemming nor lemmatization of Korean or English words, as lemmatization of Korean words still requires further work. The predictive ability of the models may be improved with more preprocessing of the text and further progress in Korean language processing.

Furthermore, even though the AUROC of the models may be acceptable in some instances, it cannot be denied that further improvements would be desirable. However, it must be noted that even between experienced triage nurses, the interrater weighted-kappa value was 0.772 [26], and 0.83 between triage nurses and experts in a retrospective review [27]. Although there is room for further improvement, even at its current state, the prediction model may be of assistance to triage personnel as an adjunct tool because expertise in triage instruments requires considerable training and may be limited in resource-constrained settings. This demand was clearly shown in a survey of triage nurses, where the most requested feature for a new tool being built was an automatic severity grade calculator for the emergency severity index, which is a widely used triage instrument [28]. Due to the inherent complexity and uncertainty involved with the triage of patients, misclassification, over-triage, and under-triage are always possible [27], and a rule-based decision support system for triage has been shown to reduce classification errors [29]. A support tool based on machine learning and NLP may also reduce triage errors and may be more robust to out-of-vocabulary terms than a rule-based system; this would be worth exploring in future studies.

There are several other possible use cases for an automated or supportive triage tool. In emergency dispatch centers and in prehospital ambulances where training and the maintenance of the quality of the larger pool of personnel may be difficult, an automatic tool may be used to support users in

the field or as a quality-control monitoring tool. An automatic tool with triage capabilities developed for in-home use may also be employed by patients for self-assessment in deciding whether to go to the emergency department or not. If automated triage tools are shown to be useful, a separate tool developed for use in disaster zones and warzones may be valuable, where personnel trained in triage would most likely be overwhelmed.

This study had several limitations. As a single center study without an external validation dataset, the results cannot be confidently generalized outside of the study hospital. Another limitation is the possibility of systemic bias in the practices of different triage nurses. Since some subjective judgement is included into the determination of the KTAS level, any systemic bias would also be reflected by the model created and further limit generalizability. More recent techniques in natural language processing, such as recurrent neural networks and convolutional neural networks for text data, which are powerful tools, were not covered in this study. These can be explored in future studies.

In conclusion, machine learning models can robustly predict the KTAS level at triage, which may have many possibilities for use, and the addition of text data improves the predictive performance compared to that achieved using structured data alone.

Conflict of Interest

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

ORCID

Sae Won Choi (<http://orcid.org/0000-0002-0123-8227>)

Taehoon Ko (<http://orcid.org/0000-0002-4045-0036>)

Ki Jeong Hong (<http://orcid.org/0000-0003-3334-817X>)

Kyung Hwan Kim (<http://orcid.org/0000-0002-2718-8758>)

References

1. National Emergency Medical Center. Emergency Medical Statistics Annual Report 2018 [Internet]. Seoul, Korea: National Emergency Medical Center; c2019 [cited at 2019 Oct 18]. Available from: https://www.e-gen.or.kr/nemc/statistics_annual_report.do.
2. Park J, Lim T. Korean Triage and Acuity Scale (KTAS). *J Korean Soc Emerg Med* 2017;28(6):547-51.
3. Ryu JH, Min MK, Lee DS, Yeom SR, Lee SH, Wang IJ,

- et al. Changes in relative importance of the 5-level triage system, Korean Triage and Acuity Scale, for the disposition of emergency patients induced by forced reduction in its level number: a multi-center registry-based retrospective cohort study. *J Korean Med Sci* 2019;34(14):e114.
4. Kwon H, Kim YJ, Jo YH, Lee JH, Lee JH, Kim J, et al. The Korean Triage and Acuity Scale: associations with admission, disposition, mortality and length of stay in the emergency department. *Int J Qual Health Care* 2019;31(6):449-55.
 5. Hinson JS, Martinez DA, Cabral S, George K, Whalen M, Hansoti B, et al. Triage performance in emergency medicine: a systematic review. *Ann Emerg Med* 2019;74(1):140-52.
 6. KTAS education guide [Internet]. Seoul, Korea: Korean Triage and Acuity Scale; c2019 [cited at 2019 Oct 17]. Available from: <http://www.ktas.org/education/info.php>.
 7. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018;46(4):547-53.
 8. Gupta A, Liu T, Shepherd S, Paiva W. Using statistical and machine learning methods to evaluate the prognostic accuracy of SIRS and qSOFA. *Healthc Inform Res* 2018;24(2):139-47.
 9. Park JH, Shin SD, Song KJ, Hong KJ, Ro YS, Choi JW, et al. Prediction of good neurological recovery after out-of-hospital cardiac arrest: a machine learning analysis. *Resuscitation* 2019;142:127-35.
 10. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. *Emerg Med Australas* 2018;30(6):870-4.
 11. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One* 2018;13(7):e0201016.
 12. Goto T, Camargo CA Jr, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open* 2019;2(1):e186937.
 13. 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.
 14. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23(1):64.
 15. 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.
 16. Sterling NW, Patzer RE, Di M, Schrage JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. *Int J Med Inform* 2019;129:184-8.
 17. McKinney W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference; 2010 Jun 28-Jul 3; Austin, TX. p. 51-56.
 18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
 19. Kim H. soynlp [Internet]. [place unknown]: github.com; 2019 [cited at 2019 Oct 16]. Available from: <https://github.com/lovit/soynlp>.
 20. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.
 21. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 2014;15(1):3133-81.
 22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA. p. 785-94.
 23. Nielsen D. Tree boosting with XGBoost: why does XGBoost win "every" machine learning competition? [master's thesis]. Trondheim, Norway: Norwegian University of Science and Technology; 2016.
 24. Passalis N, Tefas A. Entropy optimized feature-based bag-of-words representation for information retrieval. *IEEE Trans Knowl Data Eng* 2016;28(7):1664-77.
 25. Sidorov G, Velasquez F, Stamatos E, Gelbukh A, Chanona-Hernandez L. Syntactic n-grams as machine learning features for natural language processing. *Expert Syst Appl* 2014;41(3):853-60.
 26. Park JB, Lee J, Kim YJ, Lee JH, Lim TH. Reliability of Korean Triage and Acuity Scale: interrater agreement between two experienced nurses by real-time triage and analysis of influencing factors to disagreement of triage levels. *J Korean Med Sci* 2019;34(28):e189.
 27. Moon SH, Shim JL, Park KS, Park CS. Triage accuracy and causes of mistriage using the Korean Triage and Acuity Scale. *PLoS One* 2019;14(9):e0216972.
 28. Levis T, Schwartz D, Bitan Y. Triage nurses decision-support application design. *Proc Int Symp Hum Factors*

- Ergon Healthc 2018;7(1):52-5.
29. Dehghani Soufi M, Samad-Soltani T, Shams Vahdati S, Rezaei-Hachesu P. Decision support system for triage management: a hybrid approach using rule-based reasoning and fuzzy logic. Int J Med Inform 2018;114:35-44.