Explainable, Generalizable and Responsible AI Model to Triage Emergency Patients

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Triage helps to deliver the right level of emergency healthcare at the right time for the right person using the right resources. However, triage is vulnerable to mis-triage which causes delayed treatment, poor healthcare and ED overcrowding. This study, hence, aimed to develop an explainable, generalizable and responsible AI model that assists triage nurses. We identify the most important predictors, measure the order, direction, and effects of important predictors across triage levels, and quantify the minimum information required to develop a generalizable triage model.

1 Introduction

Globally, 24–28 million lives are lost annually due to conditions requiring emergency care, which causes 51% of mortality and 42% of all global disease burden (Chipendo et al., 2021). The burden of emergency care is 4.4 times higher in low and middle-income countries (LMICs) compared with high-income countries (Razzak et al., 2019). Ethiopia, for example, exhibited 1154 deaths per 100,000 people and 47,728 disability-adjusted life years per 100,000 people due to emergency conditions (Sultan et al., 2022). Triage puts patients in different categories to help emergency department professionals deliver the right level of care at the right time for the right patients using the right resources (Dong and Bullard, 2008)(Duko et al., 2019). Triage determines the location, time, and queue position of emergency patients. Triage systems, however, are vulnerable to mis-triages (Sax et al., 2022) (Tam et al., 2018), such as under-triage and over-triage. In the USA, mis-triage occurred in over 32%, of which 10% were under-triaged, and 90% were over-triaged (Sax et al., 2022). In Ethiopia, the under-triage rate is 30.7% and the over-triage rate is 21.9% (Abdelwahab et al., 2017).

Mis-triage mainly arises due to high workload and stress, inadequate training, lack of standardization, cognitive biases, overreliance on initial impressions, inadequate triage tools, crowded ED, and resource constraints (Sax et al., 2022), (Tam et al., 2018), (Abdelwahab et al., 2017), and (Duko et al., 2019). Improving the quality, access, efficiency, and administration of emergency care could lead to a 45% reduction in mortality, a 36% reduction in disability, and significant decreases in medical costs in LMICs (Ouma et al., 2018). To this end, several researchers have tried to develop an emergency triage model using machine learning (Choi et al., 2019). However, previous studies either poorly formulated the triage problem and/or used limited and/or context-specific information which affects the generalizability of the triage models. This study, therefore, aims to fill these gaps by 1) developing an explainable, generalizable, and responsible machine-learning model for ED triage and 2) providing answers to the following research questions: what are the characteristics of emergency patients in Ethiopia? Which machine learning algorithm is better for triage-level prediction? What are the most important predictors of triage at population and triage levels? Does the order, direction and effect of important predictors change across triage levels? How much information is required to make triage? To the best of our knowledge, no one has explored these research questions. Answering these

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Table 1: Performance of each machine-learning algorithm

Metrices	ML Algorithms					
	LR	NB	DT	RF	XGB	CB
Accuracy	85.1%	85.36%	92.13%	92.13%	94.31%	93.97%
Precision	87.21%	83.81%	92.32%	93.32%	94.07%	93.92%
Recall	83.62%	85.55%	92.74%	91.64%	94.31%	95.03%
F1-Score	85.29%	84.21%	92.55%	94.18%	94.18%	94.45%

questions helps to achieve SDG 3, which aims to ensure healthy lives and promote well-being for all at all ages.

2 Methods

An experimental research design was used to develop an explainable and generalizable model that facilitates responsible use of AI for emergency patient triage. The data were collected from January 1 to June 1, 2023, from a Comprehensive Specialized Hospital in Gondar, Ethiopia. The datasets consist of a total of 6020 instances with 19 features, including the target variable, about emergency patients' demographic information, chief complaints, and vital signs. Data pre-processing techniques such as data cleaning, data transformation, and feature engineering methods were applied. The missing values were handled using mode for categorical data and median for numerical data, and features were encoded as cardinal or ordinal based on their data type. We used natural language processing (NLP) tasks such as tokenization, stop word removal, and stemming to extract and organize the chief complaints, mentioned by patients, which were written in a sentence format. As the blue triage levels were excluded from the dataset, there are only four triage levels. The pre-processed data were divided into training and testing datasets following an 80/20% ratio and the training set was used to train the ML model using five-fold cross-validation. Domain area experts were employed to responsibly handle data gathering and pre-processing, feature encoding and engineering, and model validation. For example, randomly selected triage classes were initially validated by 2 nurses and approved 2 medical doctors so that existing triage biases are not propagated to the model. Six supervised ML algorithms, namely logistic regression, gaussian naive bayes, decision tree, random forest, Extreme Gradient Boosting (XGBoost), and CatBoost, were tested to develop a triage model. We used Optuna with five-fold cross-validation and f1-score to train and tune the parameters of ML algorithms. The performance of each model was evaluated using accuracy, precision, recall, and f-score. The most important features were identified using SHAP (SHapley Additive exPlanations). In addition, feature importance analysis using SHAP was used to quantify the minimum information needed to make triage-level decisions.

3 Results

Logistic Regression, Gaussian NB, Decision Tree, Random Forest, XGBoost, and Cat Boost achieved F1-score of 85.29%, 84.21%, 92.55%, 92.38%, 94.18%, and 94.45%, respectively, see Table 1. The CatBoost algorithm scored the highest performance, 20% higher than the previous study (Choi et al., 2019). The best performing model correctly classified 96% of green, 94% of yellow, 94% of orange and 96% of red triage levels, see Table 2. As a result, the cat boost model was selected as the best model for predicting the ED triage level and answering the next research questions.

The top five important features that determine the severity level of an emergency patient are modified early warning score (MEWS), mobility level, chief compliance, non-trauma, and systolic blood pressure, see Fig 1. However, this study revealed that the order, direction, and strength of features' effects vary across triage levels, see Fig 1. The MEWS and mobility level are the top two important features for all triage levels, except for the yellow triage level. Whereas oxygen saturation is the third for the green triage level, the ninth for the yellow triage level, the fourth for orange, and the seventh for the red triage level. The consciousness level (AVPU) of emergency patients is the third important feature for predicting the yellow, the eighth for orange and the fourth for red triage levels but is it not in the top 10 list of predicting features of the green triage levels.

Table 2: Classification Report

	ML Algorithms			
Triage	Precision	Recall	F1-score	Support
Green	0.92	0.97	0.95	71
Yellow	0.91	0.94	0.93	390
Orange	0.95	0.93	0.94	553
Red	0.98	0.96	0.97	181
Accuracy			0.94	1195
Macro avg	0.94	0.95	0.95	1195
Weighted avg	0.94	0.94	0.94	1195

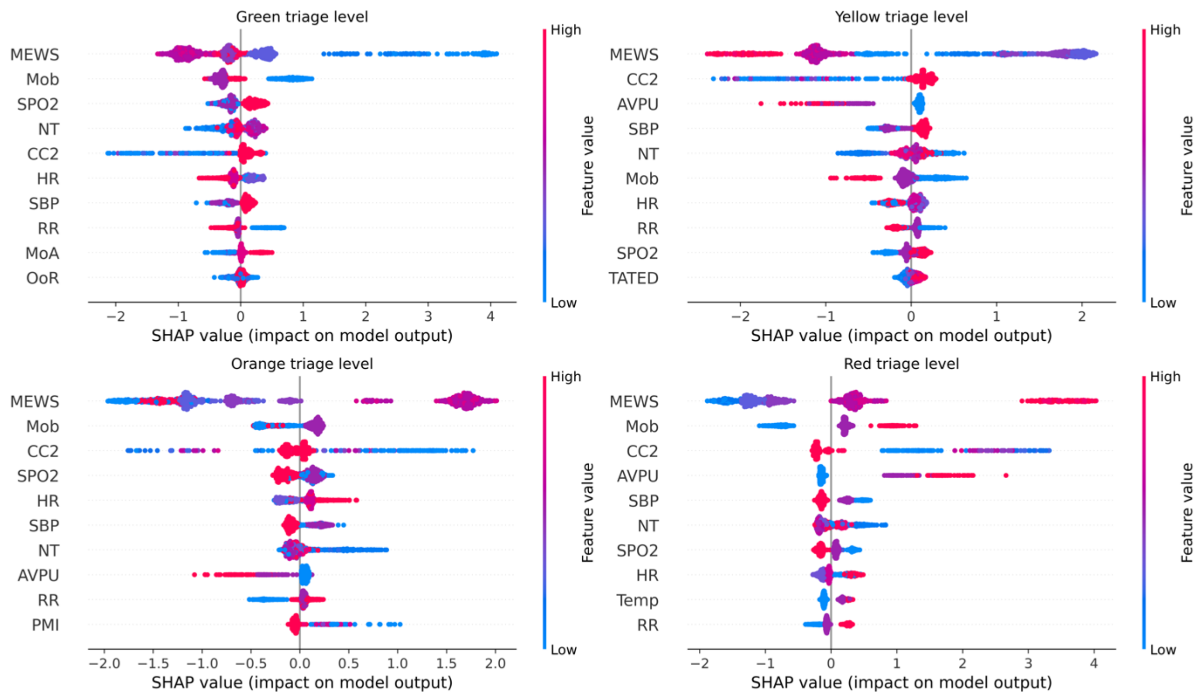

Figure 1: Top ten important features for each triage class

For AI to succeed in health, trust in the interactions between humans and ML systems is a pivotal concept which can be achieved by adding explainability features to the ML models. To this end, we used SHAP to explain our model's decisions at a patient level which increases our model's probability of being an ED team member. For example, patient number 2 was categorized as orange triage level with 99% probability and the model explains its decision in a human-understandable format, see Fig 2. The top five reasons why the model categorized patient number 2 into the orange triage level are the patient 1) has high MEWS (6), 2) arrived at ED by taxi, 3) has low oxygen saturation (<=90), 4) has a normal temperature (35.1-37.2) and 5) is alert. On the other hand, the patient's gender (male), and age (73) reduced his probability of being categorized in the orange triage level.

To quantify the minimum information required for triage, we conducted a sequence of feature importance analyses using SHAP and the best-performing model as the starting point. We first remove the last feature from the least informative features, retrain the model by tuning its parameters using Optuna with five-fold cross-validations, measure the model's performance using the test dataset and compare the new performance with the original performance. This process continues until the new performance becomes less than the original performance. As a result, without considering the pre-hospital care, mode of arrival, respiratory rate, gender, arrival time at ED, systolic blood pressure, temperature, chief complaint 1 (mentioned by the patient), the origin of reference or path to ED, heart rate, and address, similar performance (94%) can be achieved, see Table 3. However, heart

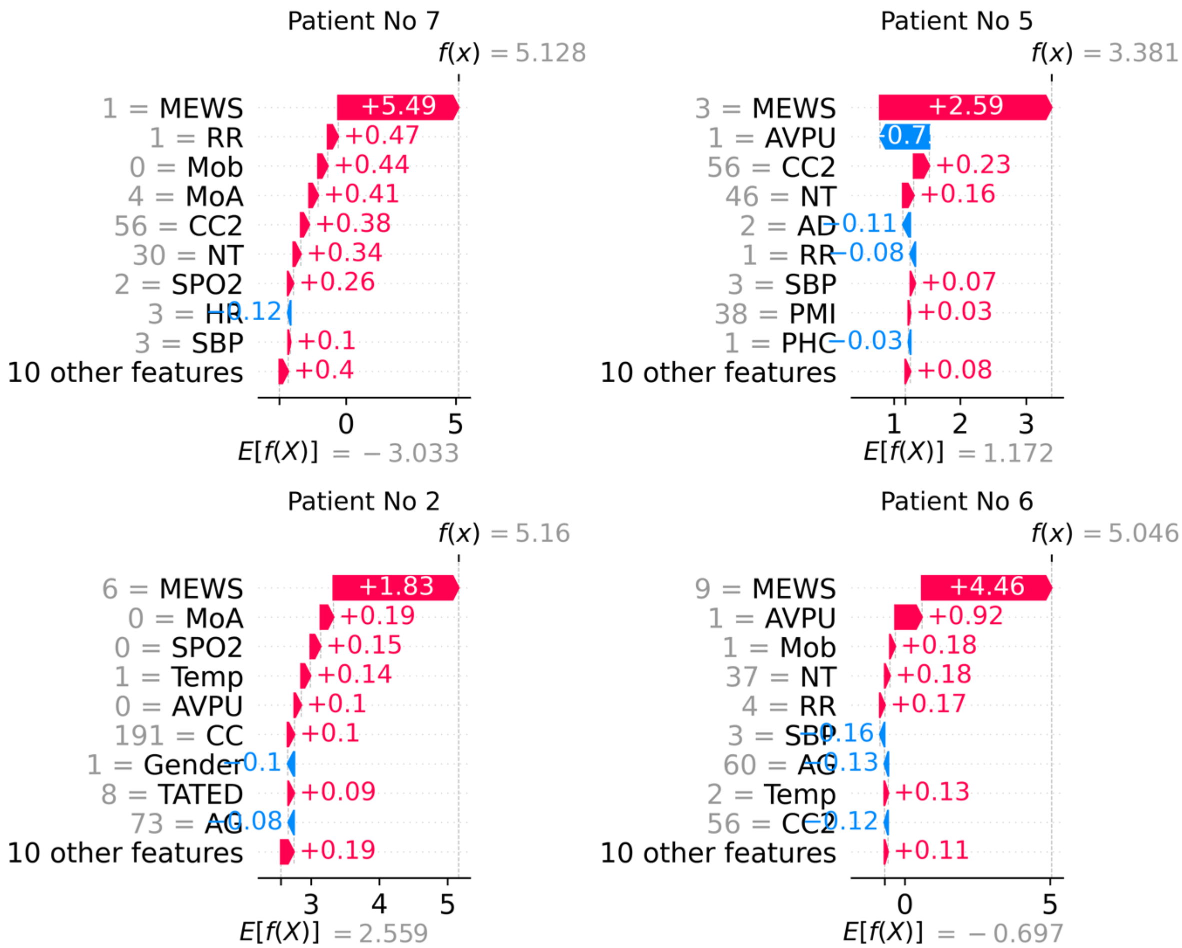

Figure 2: Top ten important features for each triage class

rate, respiratory rate, temperature, and systolic blood pressure are used to calculate the modified early warning score (MEWS). For this reason, we retained these four features and retrained the model which scored an F1 score of 94.36% on five-fold cross-validation. Hence, we can conclude that pre-hospital care, mode of arrival, gender, arrival time at ED, temperature, chief complaint 1 (mentioned by the patient), path to ED, address and previous medical illness are less important in predicting triage levels of emergency patients. In other words, triage nurses must have information about MEWS, chief complaint 2, non-trauma, mobility level, consciousness level, systolic blood pressure, oxygen saturation, age, mode of arrival, heart rate, temperature, and respiratory rate to make triage at ED. This analysis also shows that context-specific (e.g., address, path to ED) and historical (e.g., pre-hospital care and previous medical illness) data are not relevant to making ED triage decisions. In other words, our model exhibits more generalizability than the previous study (Choi et al., 2019) which 1) used context-specific data and 2) did not test their model's generalizability by removing location- and/or context-specific data.

4 Conclusion

This study aimed at developing an explainable, generalizable and responsible AI model that assists triage nurses and reduces mis-triage which in turn reduces delayed care, and ED crowding, and improves emergency healthcare outcomes. Six supervised machine learning algorithms, namely logistic regression, Gaussian NB, decision tree, random forest, XGBoost, and Cat Boost, were tested and the cat boost scored the highest performance. The top five important features are modified early warning score, mobility level, chief compliance, non-trauma, and systolic blood pressure. The order,

Table 3: Sequential feature importance analysis results

Experiment	Excluded feature	F1-score	Experiment	Excluded feature	F1-score
1	Original performance	94.45%	7	Temperature	94.25%
2	Pre-hospital care	94.49%	8	Chief complaint 1	94.31%
3	Respiratory rate	94.35%	9	Origin of reference	94.35%
4	Gender	94.37%	10	Heart rate	94.35%
5	Arrival time at ED	94.26%	11	Address	94.25%
6	Systolic blood pressure	94.22%	12	Previous medical illness	94.17%

direction, and strength of features' effects vary across triage levels and historical, location- and context-specific data are irrelevant for making ED triage decisions.

References

Abdelwahab, R., Yang, H., and Teka, H. G. (2017). A quality improvement study of the emergency centre triage in a tertiary teaching hospital in northern ethiopia. *African Journal of Emergency Medicine*, 7(4):160–166.

Chipendo, P. I., Shawar, Y. R., Shiffman, J., and Razzak, J. A. (2021). Understanding factors impacting global priority of emergency care: a qualitative policy analysis. *BMJ Global Health*, 6(12):e006681.

Choi, S. W., Ko, T., Hong, K. J., and Kim, K. H. (2019). Machine learning-based prediction of korean triage and acuity scale level in emergency department patients. *Healthcare informatics research*, 25(4):305–312.

Dong, S. L. and Bullard, M. (2008). Emergency department triage. *Evidence-Based Emergency Medicine*, pages 58–65.

Duko, B., Geja, E., Oltaye, Z., Belayneh, F., Kedir, A., and Gebire, M. (2019). Triage knowledge and skills among nurses in emergency units of specialized hospital in hawassa, ethiopia: cross sectional study. *BMC research notes*, 12:1–4.

Ouma, P. O., Maina, J., Thuranira, P. N., Macharia, P. M., Alegana, V. A., English, M., Okiro, E. A., and Snow, R. W. (2018). Access to emergency hospital care provided by the public sector in sub-saharan africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health*, 6(3):e342–e350.

Razzak, J., Usmani, M. F., and Bhutta, Z. A. (2019). Global, regional and national burden of emergency medical diseases using specific emergency disease indicators: analysis of the 2015 global burden of disease study. *BMJ global health*, 4(2):e000733.

Sax, D., Mark, D., Vinson, D., Kene, M., Ballard, D., Warton, M., and Reed, M. (2022). 133 rates and predictors of emergency department mis-triage: A multiyear, multicenter study. *Annals of Emergency Medicine*, 80(4):S62.

Sultan, M., Waganew, W., Beza, L., GebreMedihin, Y., and Kidane, M. (2022). The status of facility based emergency care in public hospitals of ethiopia using who assessment tool. *Ethiopian journal of health sciences*, 32(6).

Tam, H. L., Chung, S. F., and Lou, C. K. (2018). A review of triage accuracy and future direction. *BMC Emergency Medicine*, 18:1–7.