A Prediction Triage System for Emergency Department During Hajj Period using Machine Learning Models

Huda N. Alhazmi
hnhazmi@uqu.edu.sa
Umm Al-Qura University, College of Computers, Department of Computer Science and Artificial Intelligence, Makkah, SA

Abstract
Triage is a practice of accurately prioritizing patients in emergency department (ED) based on their medical condition to provide them with proper treatment service. The variation in triage assessment among medical staff can cause mis-triage which affect the patients negatively. Developing ED triage system based on machine learning (ML) techniques can lead to accurate and efficient triage outcomes. This study aspires to develop a triage system using machine learning techniques to predict ED triage levels using patients' information. We conducted a retrospective study using Security Forces Hospital ED data, from 2021 through 2023 during Hajj period in Saudia Arabi. Using demographics, vital signs, and chief complaints as predictors, two machine learning models were investigated, naming gradient boosted decision tree (XGB) and deep neural network (DNN). The models were trained to predict ED triage levels and their predictive performance was evaluated using area under the receiver operating characteristic curve (AUC) and confusion matrix. A total of 11,584 ED visits were collected and used in this study. XGB and DNN models exhibit high abilities in the predicting performance with AUC-ROC scores 0.85 and 0.82, respectively. Compared to the traditional approach, our proposed system demonstrated better performance and can be implemented in real-world clinical settings. Utilizing ML applications can power the triage decision-making, clinical care, and resource utilization.

Keywords:
Emergency department (ED); Triage, Critical care; Prediction; Machine learning (ML); XGB model; DNN model.

1. Introduction

The demand for a better quality of care has risen extensively. As the healthcare system develops, the development of emergency medical services (EMSs) has changed rapidly. To provide efficient care in emergency department (ED), vital decisions are required [1]. Various triage settings have been utilized in emergency department such as Canadian Emergency Department Triage and Acuity Scale [2], Australian Triage Scale [3], Manchester Triage Scale [4], and Emergency Severity Index (ESI) [5]. Despite they have been widely adapted by several EDs, they immensely rely on clinical decision which led to high variability [6]. The emergency system has been subjected to overcrowded services, resulting in increasing the patients' waiting times and treatment delay [7, 8]. The concept of triage has been defined to include severity which indicates to the essential emergency procedures, also urgency that indicates to the need of medical attention. Therefore, classifying the patients according to severity and urgency is to prioritize their treatment in the emergency department. Failure in distinguishing the risk level of patients results in under-triage or over-triage. That leads to a negative impact on the efficiency of the resource allocation and negatively affect the patients and the healthcare professionals. [9]. Therefore, triaging the patients within the emergency department is essential process.

Patient triage in ED is implemented by trained healthcare staff based on their experience and the protocols of the emergency services. In most healthcare setting, evaluating the patients and making decisions regarding the priorities falls upon the triage staff which may lead to variability in triage outcomes [10]. However, making decisions can be hindered by factors such as limited resources, dealing with many patients, and disordered setting within the emergency department [11]. Furthermore, accurate triage is vital for patient safety. As a result, efficient triage systems are necessity to facilitate timely and accurate decision-making [12].

Artificial intelligence (AI) is a computer technology can perform intelligence tasks demands human knowledge such as decision making and problem solving [13, 14]. The development of AI followed in instant development in machine learning (ML) research which can help in improving the healthcare services [15]. Considerably, ML applications in emergency medicine field can serve as a supporting tool for clinical decision-making and address the challenges in the emergency department such as triage outcomes [16]. As the number of patients increased, the need for efficient systems to evaluate patients and allocate priorities is required.

Machine learning techniques have been used as a prediction model in ED outcomes [17, 18, 19, 20].

Manuscript received July 5, 2024
Manuscript revised July 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.7.2

Recent research has demonstrated that machine learning models outperform the conventional approaches in predicting the ED triage outcomes [21, 22, 23, 24, 25, 26]. Therefore, this study explores the machine learning algorithms to predict triage outcomes within the hospital emergency department. The aim of this study is to develop a triage system based on machine learning models using patients' ED information to predict the triage levels at ED during Hajj period in Saudi Arabia. We conducted a retrospective study using 11,584 patients' visits from emergency department at Security Forces Hospital, Makkah, Saudi Arabia within three years 2021,2022, and 2023. We developed and validated the machine learning prediction system using two algorithms, including extreme gradient boosting (XGB) and deep neural network (DNN). To evaluate the performance, we used receiver operating characteristic curve (AUC-ROC) score. Both algorithms demonstrated a significant prediction ability with 0.85 and 0.82 for XGB and DNN, respectively. Our study shows that the performance of ML models using ED triage information for predicting patients' priority for care outperformed the conventional method.

This paper is organized as follows. The next section presents the background. Section 3 introduces the related work. Section 4 describes the methodology. In section 5 and 6, we show and discuss the results. The conclusion was presented in section 7.

2. Background

Triage is a formal process of prioritizing ED patients into categories according to the severity of their condition and the need for medical care and level of urgency at the time of arrival [27]. It is one of the most critical factors that guarantee the resources allocation and patients' clinical fairness. The triage system plays a significant role in organizing the emergency department [28]. It is a dynamic process rely on the emergency staff skills, they play a crucial role in a triage decision-making to provide appropriate and effective service to ED patients [29, 30, 31]. The effective triage system that distinguishes urgent cases from unurgent and allocate a waiting time based on the priority of the cases. Therefore, it is important to decide the emergency level or scale for each patient to provide the necessary care or treatment.

Several triage systems have been developed to assist prioritizing ED patients reliably and deliver appropriate health services. Triage scales are measurements used to determine the degree of urgency. These measurements have various acuity scales, ranging from three to five levels. The most common is the Australian Triage Scale (ATS), that issued by Australasian College for Emergency Physicians with 5 level according to the level of treatment acuity. Another international system was developed by Canada which is Canadian Emergency Department Triage and Acuity Scale (CTAS). The United Kingdom introduced their triage system called Manchester Triage System (MTS) and the Emergency Severity Index (ESI) adopted in the United States of America [32].

In Saudi Arabia, there is not a unified standard for triage. Various Saudi Arabia hospitals have used the Canadian Triage and Acuity Scale (CTAS) triage system. Many hospitals of ministries of health (MOH), or private use CTAS system [33]. The CTAS is an emergency department triage with five levels of severity. These levels are related to 5 scales as level 1 (Resuscitation), level 2 (Emergent), level 3 (Urgent), level 4 (Less Urgent), and level 5 (Non-urgent) [34]. This system determines the triage level beside patient complaints.

Annually, millions of Muslims from around the world visit Makkah in Saudi Arabia to do Hajj practices in specific time during the year. The number of pilgrims who are visiting the Holy city is increased every year and expected to exceed 10 million in year 2030 [35]. Accordingly, appropriate organized services including health service were needed. Since the Hajj rituals are performed in six days, the overcrowded movement over different places increase the risk of medical conditions. The Ministry of Health in Saudi Arabia provides free health service to pilgrims. Therefore, permanent and seasonal health centers are allocated and distributed within the major ritual places: Mina, Arafat, and the two Holy Mosques. During this period, EDs in different hospitals and health centers receive numerous cases of ill pilgrims involving critical conditions. Hence, they need to quickly identify high-risk patients and assign ED care efficiently.

3. Related Work

The rapid increase in emergency department visits has contributed to overcrowding and delays in providing the patients with the care they need [36, 37]. More affective approaches than traditional triage systems are needed to improve the patients' workflow in Eds, reduce the waiting time, and allocate the ED resources efficiently. With the advent of machine learning algorithms, different approaches are offered to automate the triage process [38]. In this section, we introduced the recent research that developed and evaluated different machine leaning models for ED triage outcomes. Table 1 illustrates some of these studies and their approaches.

Several studies have focused on traditional machine learning methods such as Deep Neural Network (DNN) and Extreme gradient boosting (XGB) algorithms [17, 18,19, 20]. Yoshihiko Raita et al. in [17] evaluated four ML models including logistic regression with Lasso regularization, random forest, gradient boosted decision tree, and deep neural network. They assessed the performance of the models for clinical outcomes, including admission to an intensive care unit or hospitalization. Their finding showed that ML models demonstrated high accuracy and AUC-ROC in predicting critical care and hospitalization outcomes. Another study [18] by ZhenZhen Gao et al. focused on the development and validation of Extreme Gradient Boosting (XGBoost) algorithm for triaging the patients in the emergency departments. XGBoost demonstrated high predictive accuracy with high AUC-ROC score. Comparing to traditional triage methods used by healthcare professionals, the model outperformed in identifying triages categories. The study [19] evaluated two machine learning algorithms to predict critical care outcomes for ED adult patients. They developed prediction model using extreme gradient boosting (XGB) and deep neural network (DNN) models. Their prediction outcomes defined as direct admission to ICU or in-hospital mortality. The model performance was compared with KTAS baseline model which developed using logistic regression. Their finding reveals that XGB model outperformed the baseline model with KTAS. Authors in [20] investigated three algorithms, naming, logistic regression (LR), gradient boosting (XGBoost), and deep neural networks (DNN) to predict the hospital admissions at the triage stage in emergency departments (ED). The three classifiers were trained using three sets of data: triage information, patient history, and all variables. Their results show that the predictive performance significantly improved by incorporating the patient history with the triage information.

Furthermore, many studies [39, 22, 23] have attempted to use another machine learning algorithms such as Decision Tree (DT), K-Nearest Neighbor (KNN), Association Rules (AR), Naïve Bayes (NB) and Neural Network (NN). Study [39] by Shervin Farahmand et al. developed a diagnostic accuracy of AI-based triage systems to identify the severity and cause of acute abdominal pain in emergency departments (ED). They compared their system with traditional clinical triage methods. All the AI models showed fair level of prediction, whereas Neural Network showed the best performance among them. Study [22] developed and validated machine-learning algorithms to predict high-risk emergency department (ED) revisits within a short period post-discharge. Various models were tested, but the stacked ensemble model exhibited highest AUROC and increased the prediction performance compared with other models. Authors in [23] evaluated and compared different machine learning techniques to identify the triage outcomes in ED. The models trained based on patient disposition outcomes instead of the actual triage labels. KNN, GBDT, XGBoost, and RF exhibited better performance.

Several research studies have focused on developing approaches that used machine learning models [40, 24, 25, 26]. Authors in [40] explored the performance of AI algorithms in identifying the need for critical care in patients upon arrival at ED. They used feedforward networks with softmax classifier then they compared its performance with Emergency Severity Index (ESI) and Korean Triage and Acuity. The results show that the algorithm demonstrated high predictive accuracy, with significant AUC scores. Research [24] proposed a deep learning system designed form three subsystems. Their system defines three types of prediction including triage level, hospitalization, and length of stay. The study emphasizes the importance of interpretability in deep learning models to facilitate their acceptance and use in clinical practice. Study [25] developed a triage system utilizing deep learning models to predict clinical outcomes using patients' information. Their triage system consists of RNN and CNN modules.

Table 1: The recent research in machine learning based triage models.

Study	Data size	Setting	Predictors	Performance Measures	ML models	Prediction Outcome
17	135,470 adult ED visits	National Hospital and Ambulatory Medical Care Survey (NHAMCS), USA	Demographic, vital signs, chief complaints, comorbidities, and mode of arrival	- AUC-ROC - Net benefit - Accuracy - Sensitivity - Specificity	- Lasso Regression (LR) - Random Forest (RF) - Gradient Boosted Decision Tree (XGB) - Deep Neural Network (DNN)	- Critical care (admission to ICU or in-hospital death) - Hospitalization (direct hospital admission or transfer)
18	276,164 patients	Beijing Chao-Yang Hospital	Aage, gender, patient's condition, and vital signs	AUC-ROC scores.	Extreme Gradient Boosting (XGBoost)	Four-level triage (I: critical, II: severe condition, III: emergency, and IV: general)
19	80,433 patients,	Korean National Emergency Department Information System.	Aage, gender, arrival mode, chief complaints, vital signs, time interval between arrival and onset, and level of consciousness.	AUC-ROC scores.	- Gradient Boosted Decision Tree (XGB) - Deep Neural Network (DNN)	Critical care (direct admission to ICU or in-hospital mortality)
20	560,486 patients	EDs of single hospital system (trauma centre, a community hospital-based department, and a suburban, free-standing department), USA	Demographic, chief complaint. hospital usage statistic. past medical history. outpatient medications, historical vitals, historical labs, imaging and EKG counts.	- AUC-ROC - Sensitivity - Specificity - PPV - NPV	- Logistic Regression (LR) - Gradient Boosted Decision Tree (XGB) - Deep Neural Network (DNN)	Hospital admission at the time of ED based on ESI triage.
39	215 patients	Imam Khomeini Complex Hospital, Tehran, Iran	Age, gender, vital signs, and clinical signs	- Accuracy - Sensitivity - Specificity - PPV - NPV	- Association Rules (AR) - Clustering (CL) - Logistic Regression (LR) - Decision Tree (DT) - Naïve Bayes (NB) - Neural Network (NN)	ESI-4 scores
22	6282 adult patients	National Taiwan University Hospital Hsin-Chu Branch (NTUH-HCH)	Age, sex, pre-existing diseases, diagnosis, final disposition, vital signs, chief concern, triage level, management, medication and laboratory data.	- AUC-ROC - Accuracy - Sensitivity - Specificity	- Gradient Boosted Decision Tree (XGB) - Deep Neural Network (DNN) - Random forest (RF)	High-risk ED revisit (ICU admission or died)
23	4540 patients	ED of a university hospital in Istanbul	Demographic, vital signs, chief	- Accuracy - Recall - Precision	- Logistic Regression (LR) - Decision Tree (DT)	Disposition outcomes (0: Discharge, 1: Hospitalization to

Study	Data size	Setting	Predictors	Performance Measures	ML models	Prediction Outcome
			complaints, and chronic illness.	- F1 score	- K-Nearest Neighbor (KNN) - Support Vector Machine (SVM) - Multi-layer Perception Neural Network (MLP) - Gradient Boosting Decision Trees (GBDT) - XGradient Boosting (XGB) - Adaptive Boosting (AdaBoost) - Random Forests (RF).	COVID Service, 2: Hospitalization, 3: IUC, 4: Died).
40	8,981,181 adult patients	Korean national emergency department information system	Age, sex, chief complaint, arrival time, and vital signs	- AUC-ROC - Sensitivity - Specificity - PPV - NPV - F1 scores	- Feedforwad networks with softmax classifier	Critical care with ESI and Korean Triage and Acuity System (KTAS)
24	268,716 patients	National Taiwan University Hospital (NTUH)	Demographic, vital signs, and chief complaints.	- Recall - Precision - F1 score	Proposed system comprising three Subsystems: - Triage level prediction system. - Hospitalization prediction system. - Length of stay prediction system.	- Triage level - Hospitalization - Length of stay
25	118,602 and 745,441 ED visits	National Hospital Ambulatory Medical Care Survey and Medical National Taiwan University Hospital	Age, gender, vital signs, pain index visit time, triage level, and other information	- AUC-ROC - Accuracy - Sensitivity - Specificity	- Convolutional neural network (CNN) - And recurrent neural network (RNN)	Clinical outcomes: - mortality - admission to ICU

They applied their system to predict mortality and admission to ICU. Their findings show that their mothed outperformed the conventional methods by approximately 3% to 5% higher in accuracy. In study [26], the authors focused on developing and validating a practical machine-learning algorithm to identify the possibility of fatal mis-triage. The model incorporates arrival mode, age, sex, and arrival time into the system which were considered as the top-three most important features for their model as their result showed. Also, they found that including the pulse pressure and shock index as indicators features are beneficial as emergency triage characteristics.

Expanding on these previous works, the current study implemented a machine learning models to predict the triage levels at ED based on patients' demographics, vital signs, and chief complaints information. We design a multi-class classification system to identify patients' triage levels for medical care. The predictive performance of the proposed system was examined using five triage levels to achieve the best possible outcomes. The proposed ML system can provide high level of accurate predictions which can be applicable to real situations.

4. Methodology

The study was conducted on data extracted from emergency department of Security Forces Hospital, Makkah, Saudi Arabia. The system proposed by this research is a multi-class machine learning system includes two machine learning algorithms. Figure 1 presented the research framework which includes three phases, first phase is data collection, second phase is data preprocessing, and third phase includes the model development and evaluation. The following sections present these phases in more details.

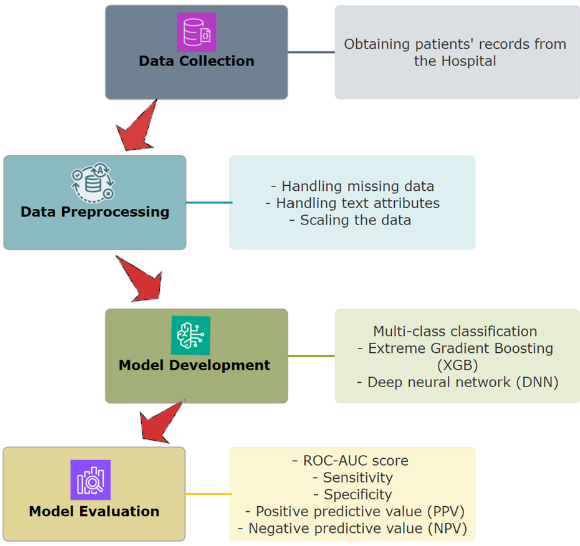

Fig 1. Research framework

4.1. Data Collection and Preprocessing

The data for this study is obtained from Security Forces Hospital, Makkah, Saudi Arabia in Hajj period within three years from 11 to 25 July 2021, 30 Jun to 18 Jul 2022, and Jun 19 to 3 July 2023. The study was approved by Biomedical Research Ethics Committee (HAPO-02-K-012-2023-10-1774), Umm Al-Qura University. The patients were triaged utilizing the Canadian Triage and Acuity Scale (CTAS) with five levels of acuity scales including resuscitation, emergent, urgent, less urgent, and non-Urgent. The collected data includes age, sex, body temperature, heart rate, respiration rate, systolic and diastolic blood pressure, blood oxygen saturation level, chief complaint, triage code, and disposition. A total of 11,584 patients' records were collected, 2426 records were excluded because of missing information, resulting in 9,158 records that used in this study as shown in Figure 2.

The dataset contains 50 attributes were grouped into categories; each category contains a set of variables. We performed some data preprocessing which includes scaling the data and handling the text attributes. Converting the categorical variables into numerical values and all the numerical variables were scaled. To convert the categorical features into numeric variables, we used the class one-hot encoder in sklearn library within Python [41]. Table 2 illustrates the variables and their description.

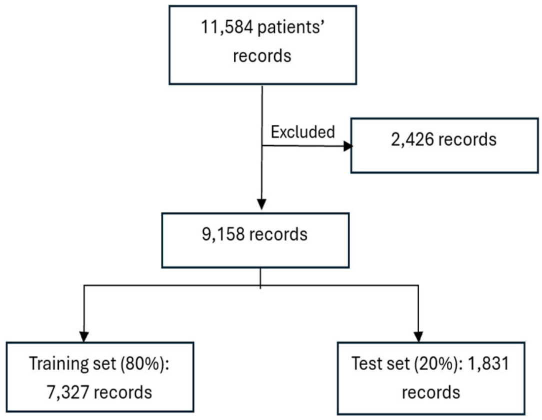

Fig 2. Study population.

4.2. Model Development and Evaluation

The prediction model is bult to solve multi-class classification problem where the classes identified by the levels from 1 to 5 where these levels represent the triage codes. We developed the system by using two machine learning algorithms called Extreme Gradient Boosting (XGB) and Deep neural network (DNN). XGB is a machine learning algorithm that belongs to the gradient boosting machine algorithms which was developed on tree-based models [42]. It has been used in regression, classification, and feature selection [43, 44]. XGB is a tree-based model that builds a set of week prediction models with minimizing the loss function. In this gradient boosting, each tree attempts to adjust the errors made by the pervious tree [45, 46, 47]. Whereas DNN is consists of multiple layers, input, output, and hidden layers. The prediction is modelled by the intermediate hidden layers, where each one consists of linear predictors that are transferred to non-linear functions [48, 49].

For machine learning predictors, we incorporated features extracted from patients' data such as age, sex, vital sign (heart rate, respiration rate, systolic blood pressure, diastolic blood pressure, body temperature, blood oxygen saturation level), and chief complaints. The data was sampled randomly with 80% (7327) as training set and 20% (1831) as a testing set, these samples were used for training and testing the XGB and DNN algorithms. For tuning the hyperparameters of each model, we selected the grid search. We implemented XGB model using 10-fold cross-validation to tune the learning rate and the maximum depth.

Table 2: Data categories and types

Category	Variable	Data type	Description
Demographics	Age	Numerical	Patient's age in term of year
	Sex	Categorical	Patient's sex converted to binary variable (1 as female, 0 as male).
Vital sign	Body temperature	Numerical	These variables represent vital sign in numbers.
	Heart rate	Numerical	
	Respiration rate	Numerical	
	Systolic blood pressure	Numerical	
	Diastolic blood pressure	Numerical	
	Blood oxygen saturation level	Numerical	
Chief complaint	The top 35 most frequent values	Categorical	Chief complaints are the symptoms that patients reported at the emergency department.
Triage code	Canadian Triage and Acuity Scale (CTAS)	Categorical	Resuscitation (1), Emergent, (2), Urgent (3), Less urgent (4), Non-urgent (5)
Response	Disposition	Categorical	This variable is represented as binary variable (1 as admission, 0 as discharge).

For DNN, we used five-layers with tuned hyperparameters such as number of hidden layers, droop-out rate, learning rate, and lambda. For implementation, we used R packages, xgboost package to build XGB model [50] and keras package for DNN model [51, 52].

The prediction system was evaluated for each algorithm, the prediction performance was calculated on the test sample with confident interval about 95%. The ROC (Receiver Operating Characteristic)-AUC (Area Under the ROC curve) score was calculated. Additionally, confusion matrix was used to calculate sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The ROC-AUC score is a performance measure indicates to the capability of the model in distinguishing the classes. The high score of this measure indicates to that the model is performing well at distinguishing the patients who were admitted to the hospital or not [53]. If the score closes to 1 means a good performance, whereas the score closes to 0 means a poor performance [54]. The confusion matrix includes the outcomes such as true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Using these values, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be calculated as shown in Equations (1,2,3,4).

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TP}{TP+FN} \quad (2)$$

$$Positive\ predicitve\ value(PPV) = \frac{TP}{TP+FP} \quad (3)$$

$$Negitive\ predicitve\ value\ (NPV) = \frac{TP}{TP+FN} \quad (4)$$

5. Result

In this study, a total of 9,158 patients were included, consisting of 5844 (64%) male and 3314 (36%) female. The triage levels of these data were (128, 1.39%) patients at level 1, (1337, 14.59%) at level 2, (4973, 54.3%) at level 3, (1731, 18.9%) at level 4, and (989, 10.7%) at level 5. Table 3 shows the results of the classification for both models. The ROC-AUC was calculated for each model as shown in Figure 3. All machine learning algorithms exhibited a significantly higher AUC. The AUC of the XGB (0.85) was higher than the value of DNN (0.82) model. Additionally, XGB model exhibited a higher sensitivity compared to DNN whereas all the models demonstrated close specificity value. Both models have high negative predictive values.

Table 3: The performance measures for each model.

Algorithm	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
GXB	0.85	0.77	0.70	0.47	0.90
DNN	0.82	0.71	0.71	0.48	0.87

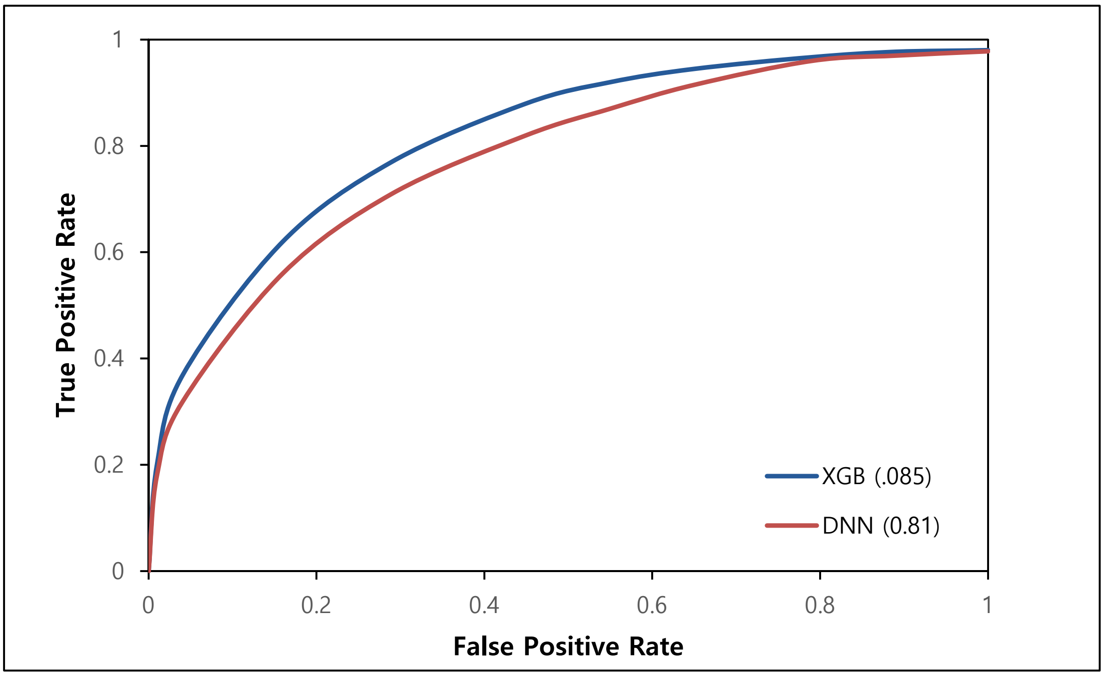

Fig 3. Receiver operating characteristic curve for the models. XGB: extreme gradient boosting and DNN: deep neural network.

The ORC diagram of the triage levels in XGB is shown in Figure 4, the AUC were calculated for all levels as 0.87, 0.83, 0.78, 0.80, 0.85, respectively. For DNN, the ORC is shown in Figure 5 and the AUC values for all levels are 0.84, 0.84, 0.75, 0.80, 0.83, respectively. Table 4 and Table 5 show the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of all levels in XGB and DNN, respectively. Level 3 and 4 demonstrated a lower sensitivity (0.73, 0.74) in XGB and (0.83, 0.82) in DNN compared to levels 1, 2, and 5 (0.86, 0.82, 0.79) in XGB and (0.83, 0.82, 0.78) in DNN, respectively. As well, for the specificity, levels in XGB (level1: 0.70, level2: 0.75, level3: 0.65, level4: 0.77, level5: 0.83) yield a higher specificity than in DNN (level1: 0.65, level2: 0.69, level3: 0.65, level4: 0.75, level5: 0.80). The sensitivity in both algorithms demonstrated close values for all levels as shown in Figure 6, whereas XGB registered a little higher specificity than DNN regarding all the levels as shown in Figure 7. In both models, most of the levels showed fair negative predictive values ranging from 0.82 to 0.91.

6. Discussion

The primary objective of ED triage is to distinguish high-risk patients accurately and prioritize the patients based on the severity of their medical cases. Improving the effectiveness of ED triaging system is a key focus of research and studies among medical and healthcare perfectional [16, 55]. Enhancing the triage process is essential for ensuring that the patients receive the actual level of care and managing the ED resources. Utilizing machine learning models in triaging process enhances the clinicals prediction operations. The main objective of this study is to accurately predict a high urgent and a less urgent level of patients at triage process in ED. We applied two machine leaning algorithms, naming, XGB and DNN incorporating variables of demographics, vital signs, and chief complaints that obtained from triage patients' records. Our study shows that the machine learning models can reliably predict the triage level of the patients visiting the emergency department.

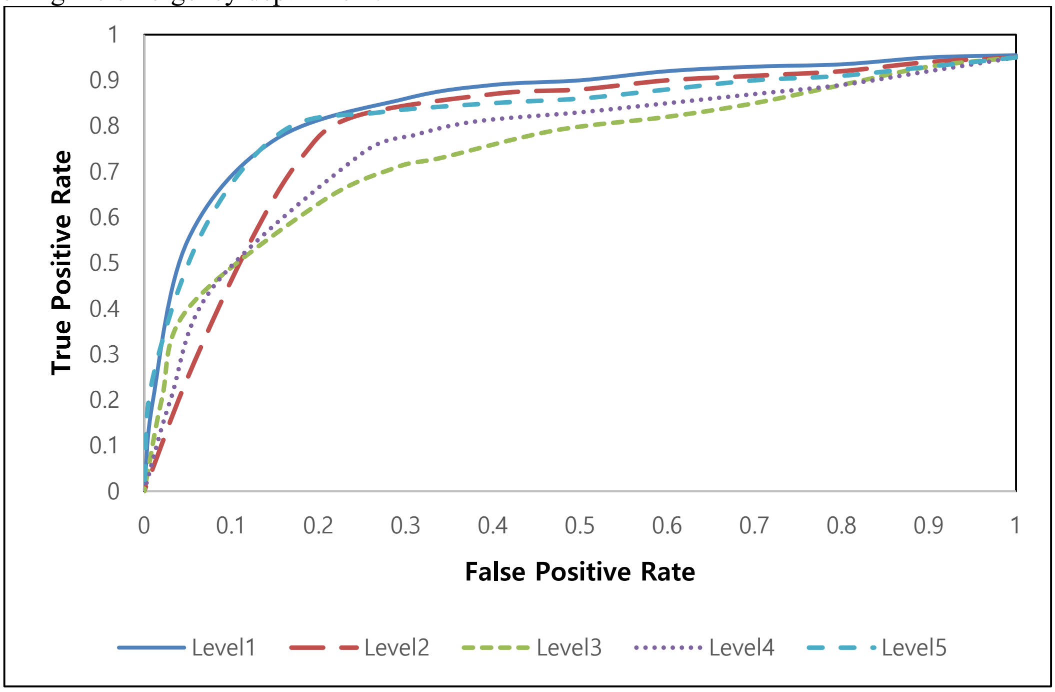

Fig. 4. XGB multi-class receiver operating characteristic curve.

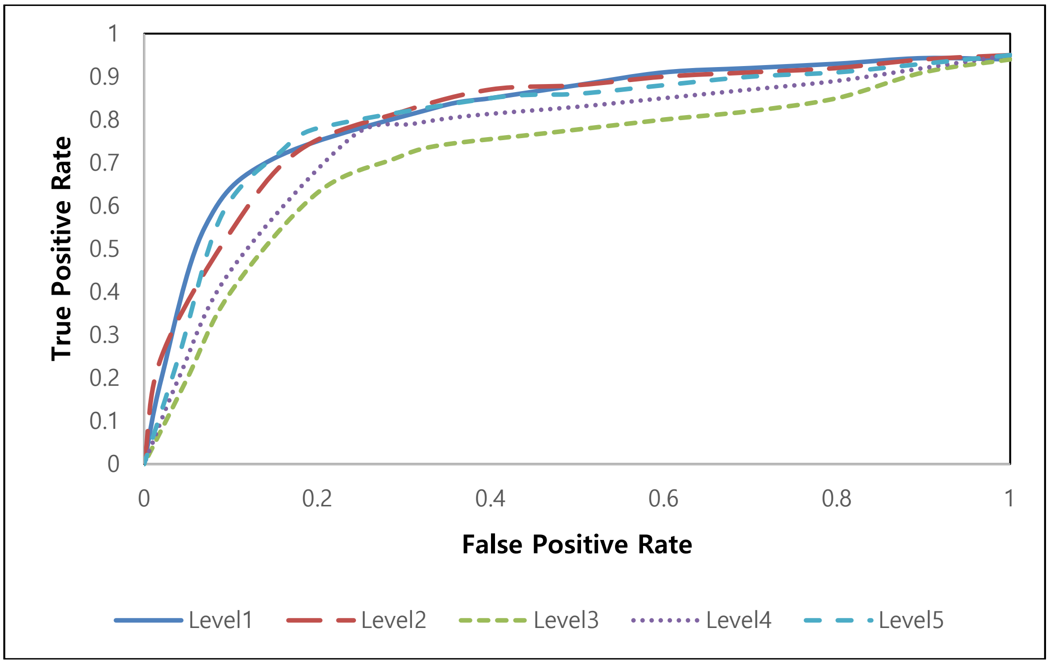

Fig 5. DNN multi-class receiver operating characteristic curve.

Table 4: The performance measures for XGB per classes.

CTAS Levels	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
Level1: Resuscitation (128, 1.39%)	0.87	0.86	0.70	0.76	0.82
Level2: Emergent (1337, 14.59%)	0.83	0.82	0.75	0.61	0.89
Level3: Urgent (4973, 54.3%)	0.78	0.73	0.65	0.35	0.90
Level4: Less Urgent (1731, 18.9%)	0.80	0.74	0.77	0.52	0.89
Level5: Non-Urgent (989, 10.79%)	0.85	0.79	0.83	0.74	0.86

Table 5: The performance measures for DNN per classes.

CTAS Levels	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
Level1: Resuscitation (128, 1.39%)	0.84	0.83	0.65	0.82	0.67
Level2: Emergent (1337, 14.59%)	0.84	0.82	0.69	0.57	0.88
Level3: Urgent (4973, 54.3%)	0.75	0.74	0.65	0.41	0.88
Level4: Less Urgent (1731, 18.9%)	0.80	0.76	0.75	0.49	0.91
Level5: Non-Urgent (989, 10.79%)	0.83	0.78	0.8	0.67	0.87

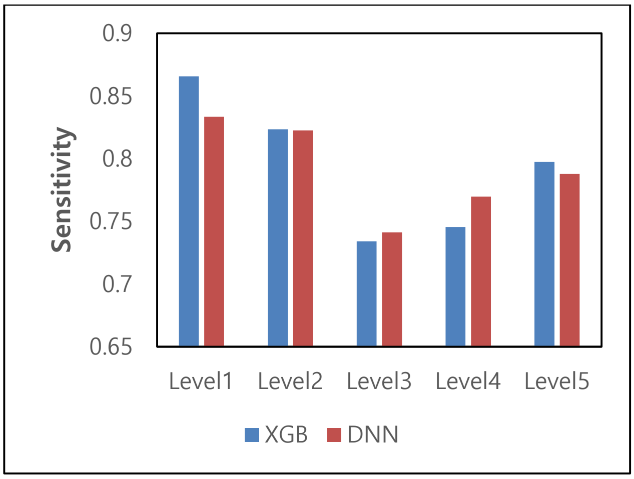

Fig 6. Sensitivity measures per classes. XGB: extreme gradient boosting and DNN: deep neural network.

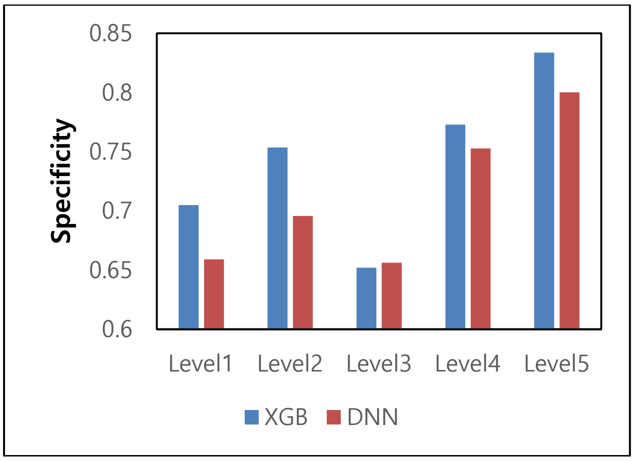

Fig 7. Specificity measures per classes. XGB: extreme gradient boosting and DNN: deep neural network.

In this study, we developed a prediction system that leverages XGB and DNN on 9,158 ED adult patients visits. The models exhibited a good prediction performance. The XGB algorithm demonstrated better distinguishing (AUC-ROC: 0.85) than DNN algorithm (AUC-ROC: 0.82). Furthermore, the prediction model including the two algorithms registered high sensitivity indicating to the decreasing of the number of under-triaged critical cases. Also, the values of the specificity for both algorithms indicate to less over-triage for the patients. Additionally, the result for triage levels revealed that both machine leaning models (XGB and DNN) demonstrated high sensitivity for Level 1 and level 2, which means that the patients who should classified as level 1 or level 2 were classified correctly. In both models, level 3 and level 4 registered very close sensitivity values but less than other levels that might means the patients who should codded as level 3 or level 4 were assigned to the lower levels. Level 5 has slightly high sensitivity value that might because less patient who needed to assign to this emergency triage code. The specificity in XGB for all levels except for level 3 are high (>70), whereas DNN exhibited lower specificity for all levels. Hence, the XGB is better than DNN in distinguishing the patients who shouldn't assigned to specific level. In term of ROC-AUC scores, all the levels in XGB and DNN registered high scores (>80) except level 3. That may suggest that the models misclassified some patients who need be classified in level 3. While our predication model demonstrated a good predictive capability, the performance needs improvement. The performance can be improved by using more predictors such as patients' history, onset times, mode of arrival, and ED resources. Our finding suggests that the machine learning models for prediction ED triage can provide an accurate triage decision and leading to improve the resource allocation and the patients' health outcomes. Furthermore, it can efficiently address over-triage and under-triage problems and enhance the patient's treatment experience.

The current study has several potential limitations. The data used in this study was obtained from a single hospital, therefore, more medical data form different medical institutes is required to improve and generalize the model. Furthermore, the machine learning algorithms were trained on data not included patient history, time of arrival, or patient mode of arrival. Expanding the data to include these features may improve the model performance.

7. Conclusion

Machine learning models provide new ways to support ED triage decision making which may enhance the patient care and the resource utilization. In ED setting, ML models showed a significant performance in predicting patients' outcomes. Using 11,584 patients' records, we developed a machine learning prediction system that can assign the patients to the correct triage levels. The system includes two machine learning models, naming XGB and DNN. These models demonstrated a superior performance in differentiation between five triage levels. Moreover, the proposed system would lower the number of under-triage and over-triage cases that caused by conventional approaches. Additionally, this study can offer support to the medical staff in precise decision making. Future work may focus on incorporating more features in the prediction model to improve the prediction performance. Moreover, combining different machine learning models in the system to provide more a broad system.

References

[1] PG. Jones, D. Mountain, and R. Forero, "Review article: Emergency department crowding measures associations with quality of care: A systematic review," Emerg Med Australas, vol. 33, no.4, pp. 592-600, 2021, doi: 10.1111/1742-6723.13743.

[2] K. Stobbe, D. Dewar, C. Thornton, et al., "Canadian Emergency Department Triage and Acuity Scale (CTAS): Rural Implementation Statement," *CJEM*,52, pp. 104-7, 2003.

[3] Australasian College for Emergency Medicine. *Guidelines on the Implementation of the Australasian Triage Scale in Emergency Department. Doc No G24*. Australasian College for Emergency Medicine; 2013:1–8.

[4] JM. Zachariasse, N. Seiger, et al., "Validity of the Manchester Triage System in emergency care: A prospective observational study," *PLoS One*, vol. 12, no. 2, 2017, doi: 10.1371/journal.pone.0170811.

[5] D. Sax, E. Warton, et al., "Evaluation of the Emergency Severity Index in US Emergency Departments for the Rate of Mistriage," *JAMA Netw Open*, vol. 6, no. 3, 2023, doi: 10.1001/jamanetworkopen.2024.23536.

[6] Y. Raita, T. Goto, M.K. Faridi, *et al.* "Emergency department triage prediction of clinical outcomes using machine learning models," *Crit Care*, vol. 23, 2019, https://doi.org/10.1186/s13054-019-2351-7.

[7] C. Morley, M. Unwin, GM. Peterson, et al., "Emergency department crowding: a systematic review of causes, consequences and solutions," PloS One, vol. 13, no. 8, 2018, doi:10.1371/journal.pone.0203316.

[8] B. Asplin, D. Magid, K. Rhodes, et al., "A conceptual model of emergency department crowding," Ann Emerg Med., vol. 42, no. 2, 2003, doi: 10.1067/mem.2003.302.

[9] A. Salmon-Rousseau, E. Piednoir, V. Cattoir, et al., "Hajj-associated infections," *Med Mal Infect*, vol. 46, pp. 346-354, 2016.

[10] B. Mistry, S. Stewart De Ramirez, G. Kelen, *et al.,* "Accuracy and reliability of emergency department triage using the emergency severity index: An international multicenter assessment," *Ann. Emerg. Med* vol. **71**, no. 5, pp. 581-587, 2018, doi: 10.1016/j.annemergmed.2017.09.036.

[11] G.umhee Baek, D. Baik, and N.Yi, "The effects of triage applying artificial intelligence on triage in the emergency department: A systematic review of prospective studies", 2023, PREPRINT available at Research Square., doi.org/10.21203/rs.3.rs-3288343/v1

[12] M. Fernandes, SM. Vieira, F. Leite, *et al*., "Clinical decision support systems for triage in the emergency department using

intelligent systems: a review," *Artif Intell Med*, vol 102, 2020, doi: 10.1016/j.artmed.2019.101762.

[13] S, Jalal, W. Parker, D. Ferguson, *et al*., "Exploring the Role of Artificial Intelligence in an Emergency and Trauma Radiology Department," *Can Assoc Radiol J.*, vol. 72, no. 1, 2021, doi: 10.1177/0846537120918338.

[14] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning, and health systems," J Glob Health, vol. 8, no. 2, 2018, doi: 10.7189/jogh.08.020303.

[15] S. Lee, N. M. Mohr, W. N. Street, *et al*., "Machine learning in relation to emergency medicine clinical and operational scenarios: an overview," *West J. Emerg. Med*, vol. 20, no. 2, pp. 219-227, 2019, doi: 10.5811/westjem.2019.1.41244.

[16] B. Mueller, T. Kinoshita, A. Peebles, *et al.* "Artificial intelligence and machine learning in emergency medicine: a narrative review," Acute *Med Surg*, vol. 9, no. 1, 2022, doi: 10.1002/ams2.740.

[17] Y. Raita, T. Goto, M. K. Faridi, *et al.* "Emergency department triage prediction of clinical outcomes using machine learning models," *Crit Care*, 2019 vol. 23, no. 1, 2019, doi: 10.1186/s13054-019-2351-7.

[18] Z. Gao, X. Qi, X. Zhang, *et al.,* "Developing and Validating an Emergency Triage Model Using Machine Learning Algorithms with Medical Big Data," *Risk Manag Healthc Policy*, vol. 15, pp.1545-1551, 2022, doi: 10.2147/RMHP.S355176.

[19] H. Yun, J. Choi, and J. H. Park, "Prediction of Critical Care Outcome for Adult Patients Presenting to Emergency Department Using Initial Triage Information: An XGBoost Algorithm Analysis," JMIR *Med Inform*, vol. 9, no. 9, 2021, doi: 10.2196/30770.

[20] 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, 2018, doi: 10.1371/journal.pone.0201016.

[21] M. Fernandes, R. Mendes, S. M. Vieira, *et al.*, "Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing," *PLoS One*, vol. 15, no. 3, 2020, doi: 10.1371/journal.pone.0229331.

[22] CW. Sung, J. Ho, CY. Fan, *et al.*, "Prediction of high-risk emergency department revisits from a machine-learning algorithm: a proof-of-concept study," *BMJ Health Care Inform*, vol. 31, no. 1, 2024, doi: 10.1136/bmjhci-2023-100859.

[23] H. Elhaj, N. Achour, M. H. Tania, *et al.*, "A comparative study of supervised machine learning approaches to predict patient triage outcomes in hospital emergency departments," *Array*, vol. 17, 2023, doi.org/10.1016/j.array.2023.100281.

[24] YT. Lin, YX. Deng, CL. Tsai, *et al.*, "Interpretable Deep Learning System for Identifying Critical Patients Through the Prediction of Triage Level, Hospitalization, and Length of Stay: Prospective Study," *JMIR Med Inform*, vol. 12, 2024, doi: 10.2196/48862.

[25] LH. Yao, KC. Leung, CL. Tsai, *et al.*, "A Novel Deep Learning-Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study," *J Med Internet Res*, vol. 12, 2021, doi: 10.2196/27008.

[26] Y. Liu, J. Gao, J. Liu, *et al.*, "Development and validation of a practical machine-learning triage algorithm for the detection of patients in need of critical care in the emergency department," *Sci Rep*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-03104-2.

[27] A. Ekwall, M. Gerdtz, and E. Manias, "The influence of patient acuity on satisfaction with emergency care: perspectives of family, friends and carers" *J Clin Nurs*, vol. 17, no, 6, 2008, doi: 10.1111/j.1365-2702.2007.02052.

[28] G. FitzGerald, GA. Jelinek, D. Scott D, *et al.*, "Emergency department triage revisited," *Emerg Med* J, vol. 27, no. 2, pp. 86–92, 2010, doi.org/10.1136/emj.2009.077081.

[29] Cioffi J. Triage decision making: educational strategies. *Accident and emergency nursing*, 1999, 7:106–11.

[30] A. Chorzempa and A. LaMotte, "The role of the triage nurse practitioner in general medical practice: an analysis of the role," *Clinical excellence for nurse practitioners*, vol. 3, no. 3, pp. 189-90, 1999.

[31] S. George, S. Read, L. Westlake, *et al.*," Nurse triage in theory and in practice," *Arch Emerg Med*, vol. 10, no. 3, pp. 220-8, 1993, doi: 10.1136/emj.10.3.220.

[32] A. M. Al Yasin, M. Alyaseen, and S. Alyaseen, "The Effectiveness of Emergency Triage Systems: A Systematic Review," vol. 6, pp. 272-282, 2023, DOI: 10.36348/sjnhc.2023.v06i08.004.

[33] M. Alquraini, E. Awad, and R. Hijazi, "Reliability of Canadian Emergency Department Triage and Acuity Scale (CTAS) in Saudi Arabia," *Int J Emerg Med*, vol. 8, no. 1, 2015, doi.org/10.1186/s12245-015-0080-5.

[34] R. Alhaqbani, R. Bahmaid, M. Almutairi, *et al.*," Patient's conception and attitude regarding triage system and waiting time at emergency department at Riyadh, Saudi Arabia," *Saudi Journal of Emergency Medicine*, vol. 3, pp. 130-137, 2022, DOI: 10.24911/SJEMED/72-1649358740.

[35] I. Kaysi, B. Alshalalfeh, M. Sayour, *et al.*," Rapid transit service in the unique context of Holy Makkah: assessing the first year of operation during the 2010 pilgrimage season," in *Proceedings of the 18th International Conference on Urban Transport and the Environment,* 2012, pp. 253-267.

[36] D. F. Gaieski, A. K. Agarwal, M. E. Mikkelsen, *et al*., "The impact of ED crowding on early interventions and mortality in patients with severe sepsis," *Am J Emerg Med*, vol. 35, no. 7, pp. 953–60, 2017, DOI: 10.1016/j.ajem.2017.01.061

[37] E. J. Weber, "Triage: making the simple complex?" *Emerg Med J*, vol. 36, no. 6, pp. 64-65, 2019, doi: 10.1136/emermed-2018-207659

[38] F. Rahimian, G. Salimi-Khorshidi, A. H. Payberah, *et al.*, "Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records," *PLoS Med*, vol. 15, no. 11, 2018, doi: 10.1371/journal.pmed.1002695

[39] S. Farahmand, O. Shabestari, M. Pakrah, *et al.*, "Artificial Intelligence-Based Triage for Patients with Acute Abdominal Pain in Emergency Department; a Diagnostic Accuracy Study," *Adv J Emerg Med*, vol. 21, no. 1, 2017, doi: 10.22114/AJEM.v1i1.11.

[40] DY. Kang, KJ. Cho, O.Kwon, *et al.*, "Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services," *Scand J Trauma Resusc Emerg Med*, vol. 28, no. 1, 2020, doi: 10.1186/s13049-020-0713-4F.

[41] Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine Learning **in** Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[42] T. Chen, C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., Association for Computing Machinery*, New York, NY, USA, 2016, 785–794.
[43] M.A. Deif, A.A.A. Solyman, M.H. Alsharif, *et al*., “Automated Triage System for Intensive Care Admissions during the COVID-19 Pandemic Using Hybrid XGBoost-AHP Approach,” *Sensors (Basel),* vol 21, no. 19, 2021, doi: 10.3390/s21196379.
[44] SK. Mun and M. Chang, “Development of prediction models for the incidence of pediatric acute otitis media using Poisson regression analysis and XGBoost,” *Environ Sci Pollut Res Int*, vol. 29, no. 13, pp. 18629-18640, 2022, doi: 10.1007/s11356-021-17135-9.
[45] A. Ogunleye and Q. -G. Wang, "XGBoost Model for Chronic Kidney Disease Diagnosis," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 2131-2140, 1 Nov.-Dec. 2020, doi: 10.1109/TCBB.2019.2911071.
[46] D. Spangler, T. Hermansson, D. Smekal D, *et al*., “A validation of machine learning-based risk scores in the prehospital setting,” *PLoS One*, vol. 14, no. 12, 2019. doi: 10.1371/journal.pone.0226518.
[47] Z. Huang, C. Hu, C. Chi C, *et al*., “An Artificial Intelligence Model for Predicting 1-Year Survival of Bone Metastases in Non-Small-Cell Lung Cancer Patients Based on XGBoost Algorithm,” *Biomed Res Int*, 2020, doi: 10.1155/2020/3462363.
[48] J. Baek, and Y. Choi, “Deep Neural Network for Predicting Ore Production by Truck-Haulage Systems in Open-Pit Mines,” *Applied Sciences*, vol. 10, 2020, doi.org/10.3390/app10051657.
[49] M. Kuhn, and K. Johnson, *Applied predictive modeling*. New York: Springer-Verlag, 2013.
[50] Xgboost, Extreme gradient boosting. https://CRAN.R-project.org/package= xgboost. Accessed: 15 Jun 2024.
[51] R Interface to Keras. https://keras.rstudio.com/. Accessed: 15 Jun 2024.
[52] D.P. Kingma and J. Ba, “Adam: A method for stochastic optimization. ArXiv14126980. http://arxiv.org/abs/1412.6980. Accessed: 15 Jun 2024.
[53] 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, pp. 1-13, 2018. doi: 10.1371/journal.pone.0201016.
[54] E. Alpaydin, Introduction to Machine Learning, 4th ed, Adaptive Computation and Machine Learning Series, MIT Press, 2020.
[55] G. Chenais, E. Lagarde, and C. Gil-Jardiné, “Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges,” *J Med Internet Res*, vol. 25, 2023, doi:10.2196/40031.