Artificial Intelligence-Based Triage for Patients with Acute Abdominal Pain in Emergency Department; a Diagnostic Accuracy Study

Shervin Farahmand[1], Omid Shabestari[2], Meghdad Pakrah[1], Hooman Hossein-Nejad[1], Mona Arbab[3], Shahram Bagheri-Hariri[4*]

1. Department of Emergency Medicine, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran.
2. Health Informatics at Institute of Health Policy, Management and Evaluation (IHPME), University of Toronto, Toronto, Canada.
3. Massachusetts General Hospital, Harvard Medical School, MA, USA.

*Corresponding author: Shahram Bagheri-Hariri; Email: hariri@sina.tums.ac.ir

Abstract

Introduction: Artificial intelligence (AI) is the development of computer systems which are capable of doing human intelligence tasks such as decision making and problem solving. AI-based tools have been used for predicting various factors in medicine including risk stratification, diagnosis and choice of treatment. AI can also be of considerable help in emergency departments, especially patients' triage.

Objective: This study was undertaken to evaluate the application of AI in patients presenting with acute abdominal pain to estimate emergency severity index version 4 (ESI-4) score without the estimate of the required resources.

Methods: A mixed-model approach was used for predicting the ESI-4 score. Seventy percent of the patient cases were used for training the models and the remaining 30% for testing the accuracy of the models. During the training phase, patients were randomly selected and were given to systems for analysis. The output, which was the level of triage, was compared with the gold standard (emergency medicine physician). During the test phase of the study, another group of randomly selected patients were evaluated by the systems and the results were then compared with the gold standard.

Results: Totally, 215 patients who were triaged by the emergency medicine specialist were enrolled in the study. Triage Levels 1 and 5 were omitted due to low number of cases. In triage Level 2, all systems showed fair level of prediction with Neural Network being the highest. In Level 3, all systems again showed fair level of prediction. However, in triage Level 4, decision tree was the only system with fair prediction.

Conclusion: The application of AI in triage of patients with acute abdominal pain resulted in a model with acceptable level of accuracy. The model works with optimized number of input variables for quick assessment.

Key words: Abdominal pain; Artificial intelligence; Emergency service, hospital; Triage

INTRODUCTION

Artificial intelligence (AI) is the development of computer systems which are capable of doing human intelligence tasks such as decision making and problem solving (1, 2). AI-based tools have been used for predicting various factors in medicine including risk stratification, diagnosis and choice of treatment. Considering the effect of uncertainty on many medical decisions, AI solutions can help in improving healthcare services (3-6). AI can also be considerably helpful in emergency departments, especially patients' triage. As numbers of patients seeking medical care have increased over the past few years, crowded emergency departments (EDs) are obliged to use an efficient system to evaluate and manage patients and allocate priorities. This structure that facilitates patient management in crowded EDs is termed triage (7-9). Applying an inappropriate triage method could result in management delays, inappropriate management and unwanted outcomes (9-11). Emergency Severity Index (ESI), which is a five-level triage system, has been applied worldwide. The ESI system is unique as it evaluates both acuity and resource utilization. This algorithm consists of five levels of care, ranging from the most to the least critical status. While Levels 1 and 2 are mainly based on high acuity level, Levels 3 to 5 emphasize resource requirements (12-15). Currently, the ESI-version 4 (ESI-4) has been applied in most hospitals in Iran. Although ESI-4 as a rule-based model should be an easy method to adopt, estimating number of required resources in

many cases is beyond the expertise of the first-line responders and requires input from an expert emergency medicine physician. If any tool could eliminate the need for this expert input, it would facilitate the assessment of patients by first-line responders and facilitate immediate communication with the nearest emergency department to arrange the requirements for quick intervention (16, 17).

To start this project, we had to choose a chief complaint to select a defined category of patients to assess with AI-based tool. Abdominal pain is one of the most common complaints so the authors chose acute abdominal pain as the selected main chief complaint. This study was done to evaluate the application of AI-based tools in patients presenting with the chief complaint of acute abdominal pain to estimate ESI-4 score without the estimate of the required resources.

METHODS

Study design

This was a prospective accuracy study conducted in Imam Khomeini Complex Hospital, Tehran, Iran, from January to March 2015. After thorough explanation of the study, written informed consent was obtained from all patients. The protocol of this study was approved by the ethical committee of Tehran University of Medical Sciences and the affiliated hospital. The data collector stripped all patient identifiable information from the data and only shared the pseudo-anonymized information with the ESI-4 assessor and the modeling team. The researchers did not interfere with the patient process management and adhered to the Helsinki Declaration principles throughout the study.

Study population

All patients older than 18 years who had visited the emergency department with acute abdominal pain as their chief complaint were eligible for the study. The participants were randomly divided and 70% of the patient cases were allocated for training the models and the remaining 30% for testing the accuracy of the models. This grouping was shared between different models to ensure comparability of the results.

Software development

We developed a web-based interface (view layer) accessible via any desktop of a smartphone (Figure 1) for using the models at the point of care. This interface allowed quick entry of the patient vital signs, location of the pain, and accompanying symptoms. The web interface passed these parameters to our prediction engine (business-logic layer). The prediction engine passed the input parameters to all models explained in the Results section. Each model provided the ESI-4 score and the perceived accuracy of the result. The prediction engine adjusted the accuracies based on the overall c-statistics score of each model and returned the result with the highest probability back to the web interface. The system user saw the ESI-4 score accompanied by the prediction probability and prioritized the cases accordingly.
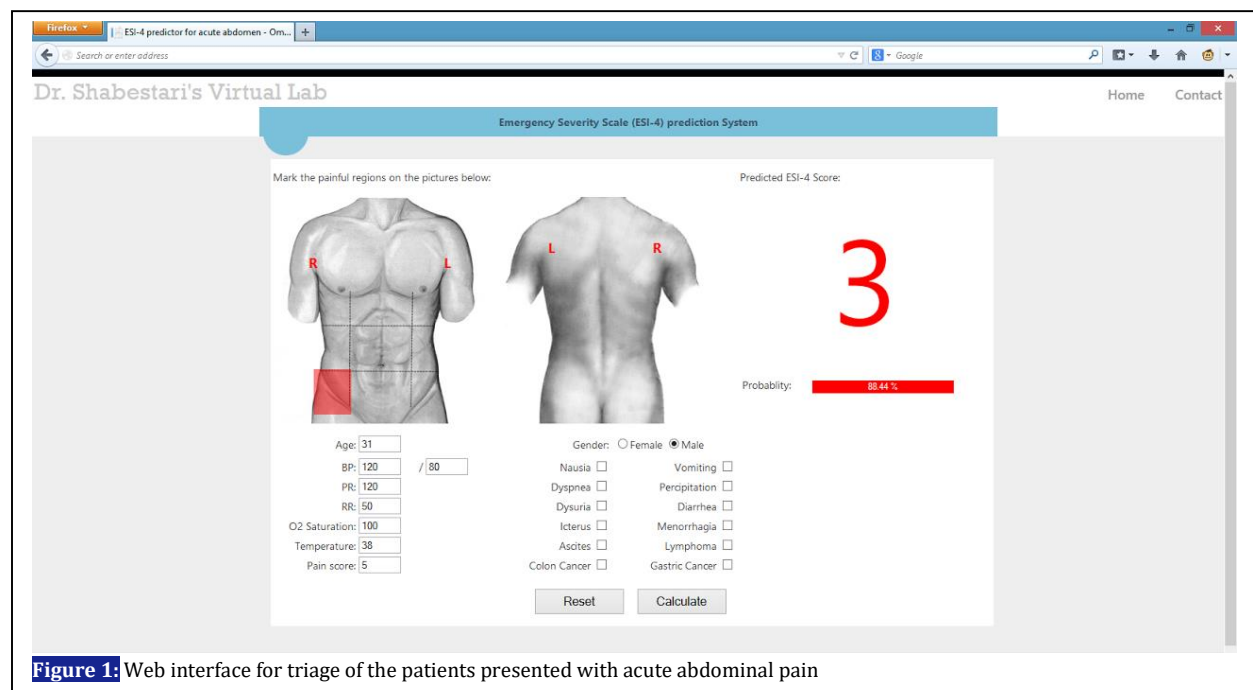

Figure 1: Web interface for triage of the patients presented with acute abdominal pain

The input variables were age, gender, vital signs (oral temperature [OT], pulse rate [PR], respiratory rate [RR], O2 Saturation [O2Sat] and blood pressure [BP]) and clinical signs (fever, precipitation, dyspnea, dysuria, diarrhea, jaundice, menorrhagia and ascites). Any reduction in input dimensions could improve the data collection time and accelerate the triage. Before training the models, we used factor analysis to reduce the input variables. A mixed-model approach including Association Rules (AR), Clustering (CL), Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB) and Neural Network (NN) algorithms, was used for predicting the ESI-4 score.

Data gathering

Demographic and baseline characteristics of patients were recorded in a pre-prepared checklist. Each patient was evaluated by a board-certified emergency medicine physician who had special training and long experience in ESI-4 scoring. The level of triage assigned by the physician was considered as the gold standard. An emergency medicine resident recruited all patients to ensure the consistency of the vital signs measurement.

Statistical analysis

Microsoft SQL Server 2014 Analysis Services was used for developing the models and IBM SPSS v.22 for calculating the c-statistics test. Fed the selected input variables plus gold standard into the algorithms during the training phase. This way the algorithms were able to learn the patterns in the data. During the test phase the models received only input variables and predicted the ESI-4 score. We compared the accuracy of the predictions between different algorithms and also evaluated the opportunity for ensemble models. During the training phase, patients were randomly selected and were given to systems for analysis. The output, which was the level of triage, was compared with the gold standard. During the test phase another group of randomly selected patients were evaluated by the system and the results were then compared with the gold standard in order to calculate accuracy, c-statistics and Kappa correlation.

RESULTS

Totally, 215 patients with a mean age of 42.13±17.83 were recruited, 115 (53%) of who were female. The mean ± SD of patients' vital signs on arrival characteristics were summarized in Table 1. The main locations of abdominal pain were right upper quadrant (35.6%), suprapubic (28.8%) and periumbilical (23.8%). One hundred fifty-one cases (70%) were enrolled in the training phase

Table 1: Vital signs of studied patients on arrival

Vital signs	Mean ± SD
Systolic blood pressure (mmHg)	123.09 ± 20.04
Diastolic blood pressure (mmHg)	74.05 ± 11.73
Pulse rate (beat/min)	86.12 ± 12.12
Respiratory (rate/min)	18.54 ± 2.42
Oral temperature (°C)	37.17 ± 0.52
O2 Saturation (%)	96.75 ± 1.98
Pain score (numerical rating scale)	5.49 ± 1.47

while the remaining patients were enrolled in the test phase. There was not a significant difference between the baseline characteristics of the patients in the training and test phases (p > 0.05).

The overall accuracy of each system was evaluated in the test phase (Table 2). In order to have a better measurement of the performance of each system, the area under curve (AUC) in the receiver operating characteristic (ROC) diagram was calculated to measure the c-statistics value of the models for each ESI-4 state. The results of these tests are available in Table 3. Levels 1 and 5 were omitted from the assessment due to the low number of cases in the testing set. This was also consistent with the number of cases in the training set. In triage Level 2, all systems showed fair level of prediction with Neural Network being the highest. In Level 3, all systems again showed fair level of prediction. However, in triage Level 4, decision tree was the only system with fair prediction.

Table 4 presents intersystem agreement using Kappa statistics. The overall agreement between the models was low, which provided the potential opportunity for developing ensemble models. The

Table 2: Overall accuracy level of different algorithms achieved in first generation models

Algorithms (Initial models)	Accuracy (%)
Association rules	70.31
Clustering	75.00
Decision Tree	73.44
Logistic Regression	68.75
Naïve Bayes	71.88
Neural Networks	70.31

Table 3: Area under the curve (AUC) in Receiver Operating Characteristic (ROC) curve for different ESI-4 states using first generation models

Algorithms (Initial models)	ESI-4 Levels		
	2	3	4
Association rules	0.734	0.722	0.418
Clustering	0.744	0.756	0.344
Decision Tree	0.714	0.660	0.713
Logistic Regression	0.761	0.713	0.257
Naïve Bayes	0.500	0.704	0.563
Neural Networks	0.769	0.685	0.219

Table 4: Kappa statistics for model agreements

Algorithms	Association rules	Clustering	Decision Tree	Logistic Regression	Naïve Bayes	Neural Networks
Association rules		0.14	0	0.54	0.58	0.55
Clustering			0	0.14	0.20	0.15
Decision Tree				0	0	0
Logistic Regression					0.56	0.83
Naïve Bayes						0.66
Neural Networks						

only exception was the excellent agreement between the Logistic Regression and the Neural Network models. This was an expected outcome as the Logistic Regression can be considered as a Neural Network model without hidden layers.

We developed meta-models by assembling the prediction results and their probabilities from the first generation of the models. Ensemble models use the outputs of the first generation models as inputs and apply different algorithms to them for predicting the outcome of interest. This technique resulted in slightly more accurate models with the Naïve Bayes algorithm performing better than all first generation models, as is presented in Table 5. Table 6 shows that the AUC of the models improved in most ESI-4 levels. This improvement was very apparent in the models that had a very low ROC value in the first generation models, such as Clustering and Logistic Regression for Level 4 of ESI-4 (Table 6). We achieved the highest overall accuracy in ensemble models with the Naïve Bayes algorithm. In the next step, we used cross-validation to check the effect of the random sampling on the accuracy of the Naïve Bayes ensemble model that had the highest accuracy among the developed models. A 10-fold

partitioning showed an average Root Mean Square Error (RMSE) of 0.18 and standard deviation of 0.03 that shows low variation in RSMEs. This shows the low risk of sampling bias in the Naïve Bayes model. We finally developed a web-based interface that can be accessed by smartphones or desktop (fig 1).

Discussion

The application of artificial intelligence in triage of patients with acute abdominal pain resulted in a model with acceptable level of accuracy. The model works with optimized number of input variables for quick assessment. One of the major benefits of designing the AI-based triage model is that it can accurately and independently triage patients in Levels 3 and 4 without estimating their resource utilization rate. This approach has two major strengths for a personalized approach compared with existing paper-based or computerized rule-based models. The first strength is the dynamic and case-based approach of the system in using different models. Although we trained our models with the same dataset, the algorithms used in the models adjust the complexity of the patterns differently. This results in varying accuracy for each case. Our operationalized interface considers this and provides the best-perceived result. The second point of strength is providing the individual accuracy for prediction to the end-user. The existing rule-based models have an overall accuracy that is based on the confidence interval (CI). The CI is an averaged measure of the overall accuracy of the model and can be variable especially in the extreme and uncommon cases. Our interface allows the triage staff to view the individual prediction probability and make informed-decisions on the probability of the accuracy.

Most patients with acute abdominal pains will consume at least one resource for diagnosis and treatment so practically they were scaled in Level 4 ESI v.4. On the other hand, almost none of the critical diagnoses for acute abdominal pain were without any other presentation or major derangement in vital signs which would obscure

Table 5: Overall accuracy level of different algorithms used in the ensemble models

Algorithms (Initial models)	Accuracy (%)
Association rules	75.00
Clustering	75.00
Decision Tree	71.88
Logistic Regression	73.44
Naïve Bayes	78.13
Neural Networks	73.44

Table 6: Area under the curve (AUC) in Receiver Operating Characteristic (ROC) curve for different ESI-4 states using ensemble models

Algorithms (Initial models)	ESI-4 Levels		
	2	3	4
Association rules	0.737	0.704	0.459
Clustering	0.790	0.739	0.751
Decision Tree	0.712	0.712	0.770
Logistic Regression	0.714	0.763	0.642
Naïve Bayes	0.635	0.708	0.839
Neural Networks	0.782	0.749	0.495

the patient from scaling in Level 1 ESI v.4. Due to the gold standard for our study, the number of patients in Level 1 and Level 5 ESI with acute abdominal pain was one. One of the major benefits of designing this AI-based triage model is that it could accurately and independently triage patients in Levels 3 and 4 without estimating their resource utilization rate. AI-based triage model could accelerate decision making in overcrowd EDs with reproducible and measurable techniques, especially in Levels 3 to 5, which is hard to estimate by triage nurses in a busy emergency room. This model can be operationalized using an easy-to-use web interface developed as a view layer that can be accessed on any desktop computer or hand-held device at the point of care. The fast response from the models can accelerate the triage process. This application can be integrated with hospital EMR systems for quick and automated entry of the input data.

Limitations

There are certain considerations in using these models. The models were trained using the cases presenting with acute abdominal pain to the emergency department. Our input cases, like most of the common abdominal pain cases, were identified at Level 2 to 4 of the ESI-4 model by our gold standard judgment. As a result, the model is mostly accurate at these ranges and will not be able to accurately predict the rare acute abdomen cases on the two extreme values of the ESI-4 categories.

CONCLUSIONS

The application of AI in triage of patients with acute abdominal pain resulted in a model with acceptable level of accuracy. The model works with optimized number of input variables for quick assessment.

AUTHORS' CONTRIBUTION

All the authors met the standards of authorship based on the recommendations of the International Committee of Medical Journal Editors.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with regard to this study.

REFERENCES

1. Ramesh A, Kambhampati C, Monson J, Drew P. Artificial intelligence in medicine. Ann R Coll Surg Engl. 2004;86(5):334-8.

2. Goletsis Y, Papaloukas C, Fotiadis D, Likas A, Michalis L. Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. IEEE Trans Biomed Eng. 2004;51(10):1717-25.

3. Mohktar MS, Redmond SJ, Antoniades NC, Rochford PD, Pretto JJ, Basilakis J, et al. Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. Artif Intell Med. 2015;63(1):51-9.

4. Houthooft R, Ruyssinck J, van der Herten J, Stijven S, Couckuyt I, Gadeyne B, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. Artif Intell Med. 2015;63(3):191-207.

5. Kuo R, Huang M, Cheng W, Lin C, Wu Y. Application of a two-stage fuzzy neural network to a prostate cancer prognosis system. Artif Intell Med. 2015;63(2):119-33.

6. Liu N, Holcomb J, Wade C, Darrah M, Salinas J. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. Shock. 2014;42(2):108-14.

7. Durani Y, Brecher D, Walmsley D, Attia MW, Loiselle JM. The Emergency Severity Index version 4: reliability in pediatric patients. Pediatr Emerg Care. 2009;25(11):751-3.

8. Fernandes C, Tanabe P, Gilboy N, Johnson L, McNair R, Rosenau A, et al. Five-level triage: a report from the ACEP/ENA Five-level Triage Task Force. J Emerg Nurs. 2005;31(1):39-50.

9. Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. Dtsch Arztebl Int. 2010;107(50):892-8.

10. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Acad Emerg Med. 2000;7(3):236-42.

11. Yurkova I, Wolf L. Under-triage as a significant factor affecting transfer time between the emergency department and the intensive care unit. J Emerg Nurs. 2011;37(5):491-6.

12. Gilboy N, Tanabe P, Travers DA. The Emergency Severity Index Version 4: changes to ESI level 1 and pediatric fever criteria. J Emerg Nurs. 2005;31(4):357-62.

13. Yoon P, Steiner I, Reinhardt G. Analysis of factors influencing length of stay in the emergency department. Cjem. 2003;5(3):155-61.

14. Fatovich D, Hirsch R. Entry overload, emergency department overcrowding, and ambulance bypass. Emerg Med J. 2003;20(5):406-9.

15. Shelton R. The emergency severity index 5-level triage system. Dimens Crit Care Nurs. 2009;28(1):9-12.

16. Abdoos M, Seyed Hosseini Davarani H, Hosseini Nejad H. Impact of Training on Performance of Triage: A Comparative Study in Tehran Emergency Department. Int J Hos Res. 2016;5(4):122-5.

17. Hossein-Nejad H, Banaie M, Seyedhosseini-Davarani S, Khazaeipour Z. Evaluation of the Significance of Vital Signs in the Up-Triage of Patients Visiting Emergency Department from Emergency Severity Index Level 3 to 2. Acta Med Iran. 2016;54(6):366-9.