Obstet Gynecol Res 2022; 5 (2): 136-157 DOI: 10.26502/ogr085

Research Article

Predictive Models for Emergency Department Triage using Machine Learning: A Systematic Review

Fei Gao[1*], Baptiste Boukebous[3,4], Mario Pozzar[1,2], Enora Alaoui[1,2], Batourou Sano[1,2], Sahar Bayat-Makoei[1]

[1]University of Rennes, EHESP, CNRS, Inserm, Research on Health Services and Management, Rennes, France

[2]University of Rennes, Rennes, France

[3]ECAMO, UMR1153, Centre of Research in Epidemiology and StatisticS, INSERM, Paris, France

[4]Hoptial Bichat/Beaujon, APHP, Paris, France

[*]**Corresponding Author:** Fei GAO, EHESP School of Public Health, Department of Quantitative Methods for Public Health, Avenue of Professor Léon Bernard, 35043 Rennes, France

Received: 22 April 2022; **Accepted:** 29 April 2022; **Published:** 26 May 2022

Citation: Fei Gao, Baptiste Boukebous, Mario Pozzar, Enora Alaoui, Batourou Sano, Sahar Bayat-Makoei. Predictive Models for Emergency Department Triage using Machine Learning: A Systematic Review. Obstetrics and Gynecology Research 5 (2022): 136-157.

Abstract

Background: Recently, many research groups have tried to develop emergency department triage decision support systems based on big volumes of historical clinical data to differentiate and prioritize patients. Machine learning models might improve the predictive capacity of emergency department triage systems. The aim of this review was to assess the performance of recently described machine learning models for patient triage in emergency departments, and to identify future challenges.

Methods: Four databases (ScienceDirect, PubMed, Google Scholar and Springer) were searched using key words identified in the research questions. To focus on the latest studies on the subject, the most cited papers between 2018 and October 2021 were selected. Only works with hospital admission and critical illness as outcomes were included in the analysis.

Results: Twenty-one articles concerned the two outcomes (hospital admission and critical illness) and

developed 75 predictive models. Random Forest and Logistic Regression were the most commonly used prediction algorithms, and the receiver operating characteristic-area under the curve (ROC-AUC) the most frequently used metric to assess the algorithm prediction performance. Boosting, Random Forest and Logistic Regression were the most discriminant models according to the selected studies.

Conclusions: Machine learning-based triage systems could improve decision-making in emergency departments, thus leading to better patients' outcomes. However, there is still scope for improvement concerning the prediction performance and explicability of ML models.

Keywords: Triage; Emergency Department/Emergency Room; Machine Learning; Modeling; Model; Classification; Predictive; Artificially Intelligence; Decision Support Systems; Patient Prioritization; Natural Language Processing

1. Background

Emergency service use has increased by approximately 35% over the last 20 years, whereas the number of emergency departments (ED) has declined by 11% in the same period [1-2]. Consequently, overcrowded emergency rooms and uneven distribution of resources are now crucial public health issues. ED, where diagnostic and therapeutic interventions must be executed rapidly and effectively [3], are one of the biggest sources of hospitalization [4-5]. On arrival at the ED, patients are first classified according to the severity of their condition, in order to provide rapid treatment to those requiring immediate medical intervention. The aim of this triage, usually performed by a nurse on the basis of the patients' vital signs and main complaint [6-7], is to optimize the waiting time and prioritize resource usage. Under-triage (low sensitivity) of critical patients can delay treatment and increase mortality. Over-triage (low specificity) may worsen emergency room overcrowding and increase waiting times, preventable costs, and resource consumption. In this context, the classification protocol used by nurses is a widely accepted tool to identify priority patients. The Emergency Severity Index (ESI), Canadian Triage and Acuity Scale (CTAS), the Manchester Triage System (MTS) and French Clinical Classification of Emergency Patients (CCMU) are some examples of triage protocols frequently used worldwide [8-10].

Recently, there has been increased interest in developing ED triage decision support systems based on big volumes of historical clinical data to differentiate and prioritize patients. Artificial intelligence (AI) and Machine learning (ML) provide a novel research area and have shown promising results. Conventional statistical models are based on pre-programmed rules derived from specific clinical predicators. On the other hand, ML prediction modeling use non-parametric algorithms that can incorporate a larger variety of complex predictors and maintain predictive performance. One of the most important advantages of ML-based methods is that they can identify patterns in the input data using algorithms, and then exploit the uncovered patterns to predict future data. It has been shown that ML can enhance the triage capacity [3, 11-15] in order to rapidly screen patients and ensure their timely treatment in function of their condition. ML-based methods can reduce human errors, time and costs, and improve the quality of care services. Various studies have analyzed the impact of ML on ED triage with different prediction objectives, such as patients' prioritization, critical illness, mortality, ED re-admission, ICU admission, ED length of stay, cardiac arrest and bacteremia, using the information available at triage [16-22]. The authors implemented one or more ML models and evaluated their prediction performance, in comparison with the conventional clinical prediction standards (e.g. ESI).

Other studies focused not only on the use of structured data, but also of the textual data generated (e.g. free textual notes by nurses or clinicians). The combination of Natural Language Processing (NLP) techniques for clinical notes and ML may further improve the predictive performance [16, 23–25]. Indeed, NLP, one of the main AI components, allows converting unstructured data, such as ED triage notes, into a set of quantitative parameters then can be used in ML-based systems [26]. The aim of this review was to assess the performance of recently described ML models used for patient triage in ED. This scoping review presents the variables, intelligent techniques, and performance measures used to develop these models and to evaluate their triage performance. We also compared the effect of including unstructured data in addition to conventional structured predictors. Lastly, we wanted to identify the future challenges of incorporating ML models in the ED workflow.

2. Method

2.1 Information sources and search strategy

The findings were reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [27]. Four databases (ScienceDirect, PubMed, Google Scholar and SpringerLink) were manually searched using key words ("triage", "emergency department"/"emergncy room", "machine learning", "modeling", "model", "classification", "predictive", "artificially intelligence", "decision support systems", "patient prioritization") identified in the research questions , as done in previous studies [28-33]. Only studies published between 2018 and October 2021 were selected to assess recent developments in this thematic area.

2.2 Selection process and eligibility criteria

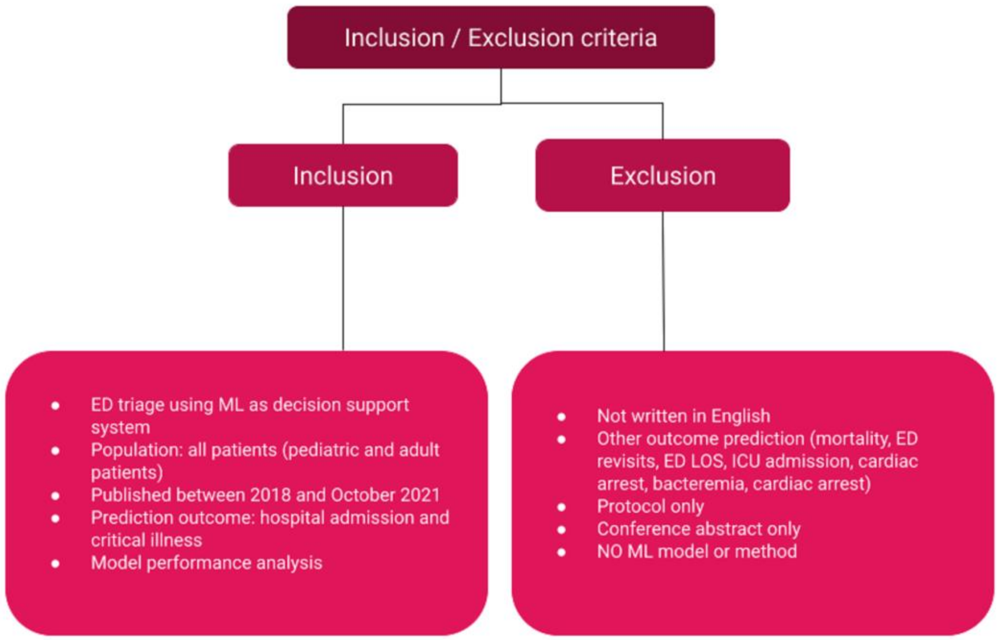
Inclusion and exclusion criteria were defined to select relevant articles (Figure 1). Based on the abstract analysis, articles were selected if i) triage has taken place in the framework of emergency care; ii) ML was used to make predictions and the ML model performance was compared using evaluation metrics. Only studies that compared the different ML models, or one or more ML models with at least one standard triage method (e.g. clinician judgment, triage-based index score) were selected; iii) the prediction outcome included hospital admission and critical illness; other outcomes, such as mortality prediction, ED re-admission, ICU admission, ED length of stay were excluded.

2.3 Data collection process

A checklist based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) was established [34], and was completed with other information (e.g. programming language and modeling libraries).

2.4 Risk of bias assessment

The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias and the applicability of the included ML models [35]. This tool uses twenty signaling questions to assess four domains, and an overall judgement on the risk of bias and applicability.



Abbreviations: ML Machine Learning, ICU Intensive Care Unit, LOS Length of Stay

Figure 1: Selection Criteria.

3. Results

3.1 Study selection

Among the 2,010 unique articles identified using our electronic search strategy, 37 were identified for full-text review (Figure 2). Twenty-one articles published between 2018 and October 2021 were selected and analyzed in detail. Their characteristics are summarized in Table 1: seven studies were performed in the USA (one included data from USA and Portugal), three in Korea, two in the Netherlands, two in Australia, one in Northern Ireland, one in Israel, one in India, one in Brazil, one in Taiwan, one in Chile and Spain, and one in Ukraine and Canada.

3.2 Data sample and predictors

In all selected articles, the study population concerned patients visiting the ED, with the exception of the article by Kim et al. that focused on the prehospital environment [41]. Sample sizes varied from ~20,000 to ~3,000,000 individuals. Only the study by Olivia et al. [45] did not provide clear information on the sample size. Figure 3 summarizes the variables used to build the ML models in each study. Hong et al. [40] included 972 explanatory variables, Roquette et al. included 62 [10] and van Rein et al. 48 [48], while the other articles used fewer than 20 predictors. Although the data used for the ML model implementation were specific to each study, several common categories could be identified, such as demographic variables (age and sex), clinical variables (vital signs and diagnosis), arrival information (time and transport mode), ED visit outcome (hospital admission or discharge).

Most authors (17/21 articles) took into account a standard triage classification index, such as ESI. Half of the articles (11/21 articles) linked data to the common main complaints, and only the studies by Goto et al., Klug et al. and Chen et al. [16, 42, 50] included information on comorbidities. Less than half of the articles presented information on the use of hospital metrics (e.g. number of previous ED visits and number of previous hospitalizations). Hong et al. [40], Rendell et al. [47], Chen et al. [50], Roquette et al. [10] and Levin et al. [44] included the patients' past medical history. Hong et al. [40], Roquette et al. [10] and De Hond et al. [37] added also information on historical laboratory test results, and imaging and electrocardiogram exams. Six articles investigated the role of NLP techniques to improve ED triage performance by including free textual triage notes. Sterling et al. [23] and Tahayori et al. [26] predicted disposition from the ED using only the triage text, while Chen et al. [50], Roquette et al. [10] and Choi et al. [13] combined structured predictors and unstructured textual triage notes. Ivanov et al. [51] used information collected at triage in addition to the textual patient history to predict the patient priority-zation score, and compared the performance of their model to the ESI-based triage performed by nurses.

3.3 Machine learning process

3.3.1 Candidate variable handling and feature engineering: In the majority of the selected studies, all variables were included in the implemented models (Figure 4). Rendell et al. [47], Kwon et al. [43], Fernandes et al. [38] and Araz et al. [36] used Stepwise or Correlation-based methods for feature selection to reduce the number of input variables. When building a predictive model, it is often possible to improve its predictive performance by transforming variables. The most common transformation methods include categ-orization (e.g. bucketing, binning), interactions, and polynomial or spline transformation for numerical variables. Only Rendell et al. proposed predictor interaction features [47]. None of the authors used polynomial or spline transformation. Levin et al. [44] and Kim et al. [41] did not provide any clear information on the variables retained in their models.

3.3.2 Data resampling: In most articles, the datasets were randomly partitioned into training and test datasets (Table 1). The percentage of data contained in each dataset differed among studies (e.g. 90:10 in the study by Hong et al. [40], and 70:30 in the study by Raita et al. [46]). Levin et al. used the bootstrapping resampling technique [44]. Fourteen studies used the cross-validation method to validate the model perfor-mance or to tune hyperparameters, which helps to avoid the risk of overfitting or underfitting [52-53].

3.3.3 Prediction algorithms and calibration of hyperparameters: In total, 75 models were used to predict hospital admission or critical illness outcomes (Figure 5). Deep Neural Network and Logistic Regression (n=12/21 articles) were the two most widely used models, followed by Random Forest (n=11/21 articles) and Gradient Boosting models (n=10/21 studies). Some models were only used in one study: K-Nearest Neighbors (Rendell et al. [47]) and Random Under Sampling Boost (Fernandes et al. [38]). Among the used tools, R and Python were the most common, followed by Java, MATLAB and the SQL language. Only 11/21 articles included infor-mation on calibration of at least one hyperparameter, depending on the method used [10, 16, 23, 36-40, 46, 49, 51].

3.3.4 Evaluation metrics: The metrics used to evaluate the performance of the different models (Figure 6) included the F1 score, the receiver operating characteristic-area under the curve (ROC-AUC), sensitivity and specificity, and accuracy. The sensitivity and specificity and ROC-AUC metrics were the most used.

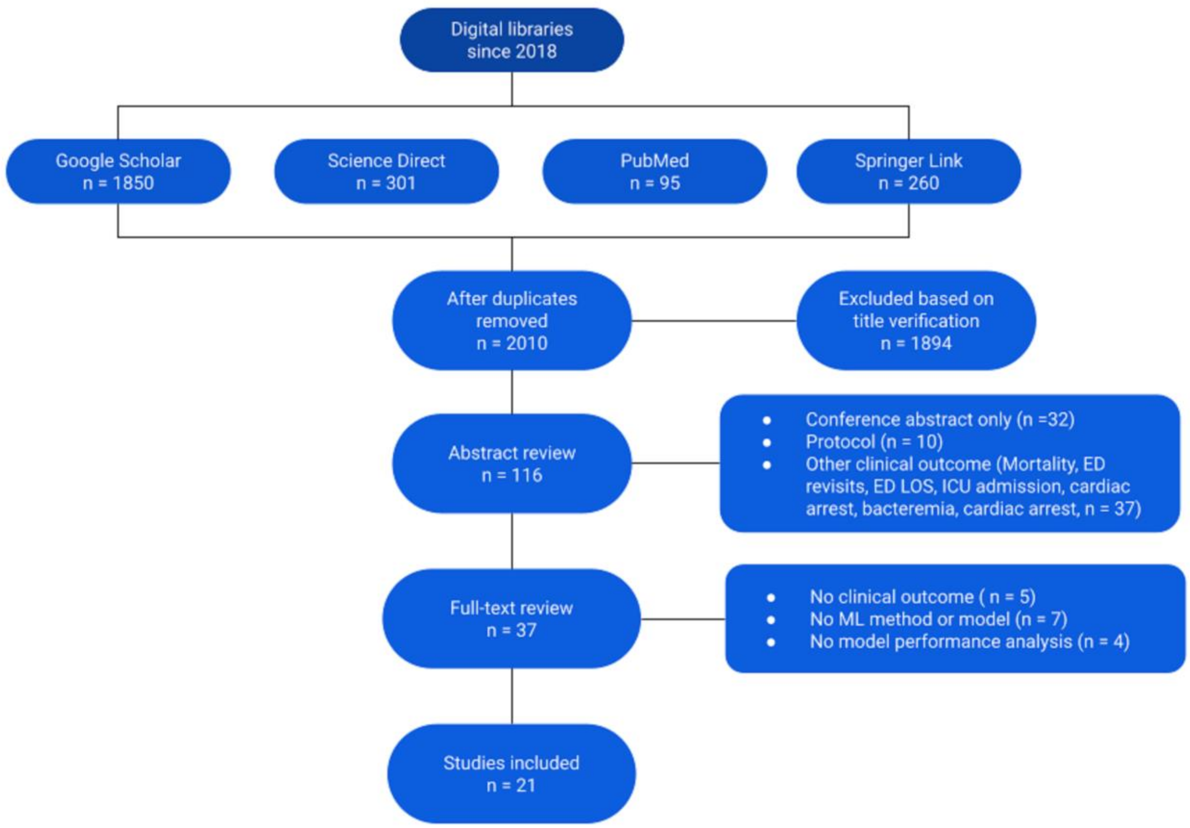
3.3.5 Model agnostic methods: Most authors used Logistic Regression coefficients to identify significant variables. For models that cannot be interpreted directly, such as Random Forests, Gradient Boosting and Neural Networks, the Permutation Feature Importance model-agnostic method was used in eight studies to identify the variables that most contributed to discrimination [16, 37-38, 40-42, 44, 46]. This method assesses the predictor importance by measuring the increase of the prediction error when the feature values are permuted.

3.4 Model performance assessment

3.4.1 Hospitalization outcome: In the selected studies, 60 models were developed (Table 1) with hospital admission as outcome. Figure 7 illustrates the performance of the prediction models based on the C-statistic method (AUC). Gradient Boosting was the most discriminant (median AUC = 0.863 and interquartile ranges (IQR) = 0.859-0.891), compared with Logistic Regression (median AUC = 0.845, IQR = 0.806-0.887), Lasso-ElasticNet (median AUC = 0.825, IQR = 0.818-0.833) and Single Layer Neural Networks (median AUC = 0.825, IQR = 0.820-0.830), and also Deep Neural Networks and K-Nearest Neighbors (median AUC = 0.822 for both, IQR = 0.795-0.876 and 0.815-0.850, respectively).

3.4.2 Critical illness: Fifteen models used critical illness as outcome measure (Figure 8). Deep Neural Networks displayed the best performance in differentiating between patients with and without a critical illness (median AUC = 0.875, IQR = 0.857-0.895), followed by Random Forest (median AUC = 0.870, IQR = 0.850-0.881), Logistic Regression (median AUC = 0.851, IQR = 0.846-0.860), and Gradient Boosting (median AUC = 0.840, only one model).

3.4.3 Natural language processing: Bag-of-words (BOW), term frequency-inverse document frequency (TF-IDF), word embedding [54], and paragraph vectors [55] were common NLP techniques used to transform free text into numerical variables [50]. Serling et al. [23] and Tahayori et al. [26] used NLP to process the free text of the nurses' triage notes, independently of other clinical variables, to predict patient outcomes. Serling et al. [23] tested three common NLP techniques: BOW, word embedding, and paragraph vectors. Model training and validation were performed using Deep Learning for Java (DL4J). They found that the best prediction for hospital admission were obtained with paragraph vectors (AUC = 0.785, 95% CI = 0.782 - 0.788). Tahayori et al. [26] used the Bidirectional Encoder Representations from Transformers model and obtained an AUC of 0.88 for predicting hospital admission. Ivanov et al. used the C-NLP method from the OpenNLP Java library to extract medical historical data. They found that when using the NLP information, accuracy of predicting ED disposition was 26.9% higher than the mean nurse accuracy (75.7% versus 59.8%) (P <.001). Chen et al. compared the prediction performance of only structured predictors, of only clinical narratives recorded following the Subjective, Objective, Assessment, and Plan (SOAP) method, and of the combination of structured data and unstructured texts. They showed that the third option gave the best prediction metric. Figure 9 shows the performance of the different models with and without NLP techniques. Roquette et al and Choi et al did the same analysis using free triage notes instead of SOAP notes, and concluded that the addition of nursing triage text data improved the prediction performance of all the studied models.



Abbreviations: ML Machine Learning, ICU Intensive Care Unit, LOS Length of Stay

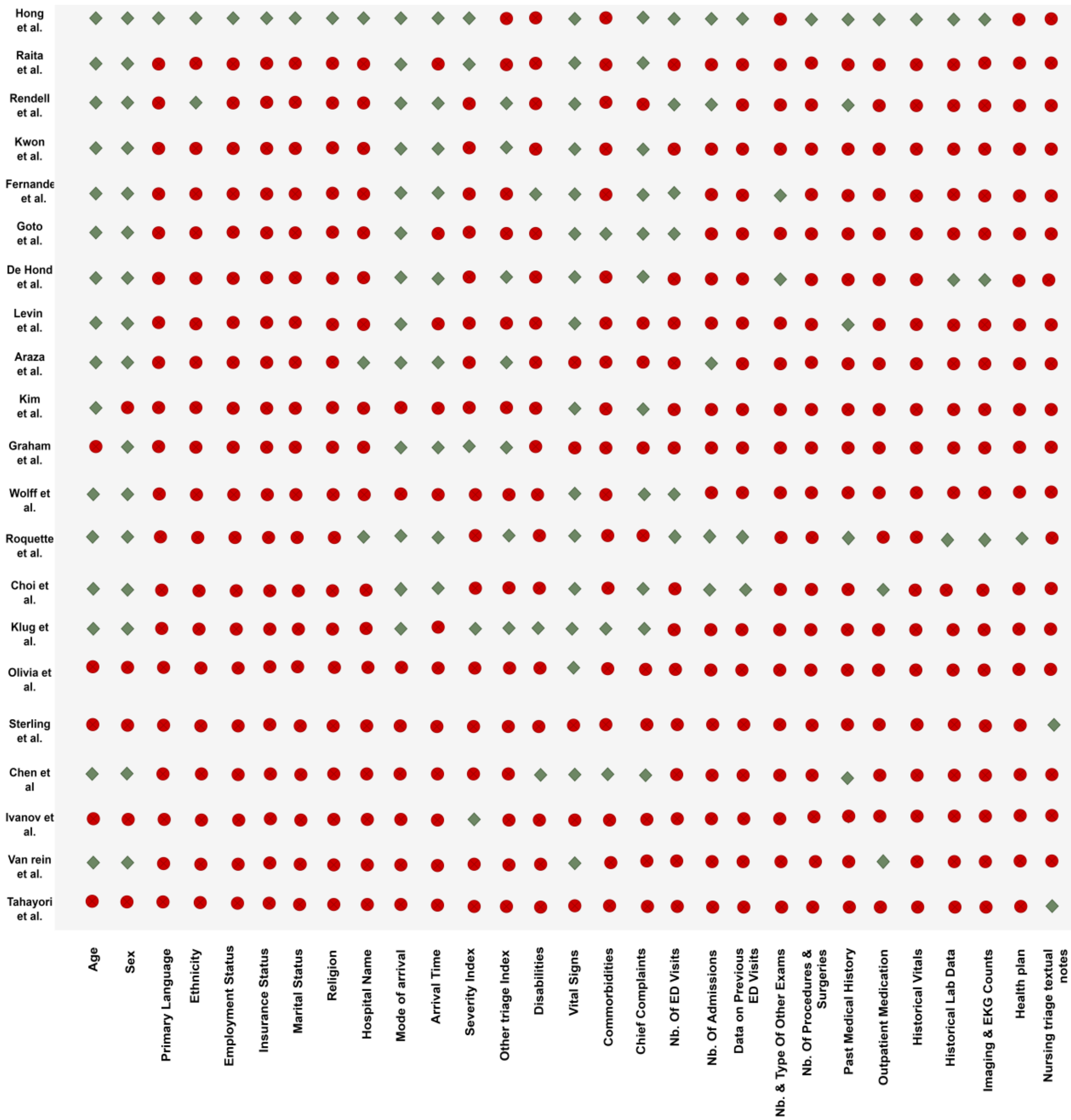
Figure 2: Flow diagram of database search and final selection.

Author	Year	Country	Population	Outcome	Method Used	Predictors	Sample size	Validation Method	Tools (R, Python packages)
Araz et al. [36]	2019	USA	ED	Hospitalization	LR, ANN, DT, RF, SVM, XGBoost	7	118,005	Randomly partitioned (70: 15:15) Training: validation: test	R (packages: glm, NeuralNetTools, ksvm, randomForest, XGBoost)
Choi et al. [13]	2019	Korea	ED	KTAS level	LR, RF, XGBoost	10	138,022		Python (pandas, scikit-learn, soynlp libraries)
De Hond et al. [37]	2021	Netherlands	ED	Hospitalization	LR, RF,GBDT, DNN	20	172,104	Split sample (66.6:33.3) + cross validation	Python, R
Fernandes et al. [38]	2019	Portugal and USA	ED	Hospitalization (ICU)	LR, RUSB, RF	13	599,276 and 267,257 respectively	Random stratified sample (70:30) + 10-fold cross - validation	
Goto et al. [16]	2019	USA	ED	Critical illness and hospitalization	LASSO, RF, GBDT, DNN	8	52,037	Random split sample (70:30) + cross-validation	R, Keras
Graham et al. [39]	2018	Northern Ireland	ED	Hospitalization	LR, RF,GBDT	13	107,545	Split sample (80:20) + cross-validation	SQL, R
Hong et al. [40]	2018	USA	ED	Hospitalization	LR, GBDT, DNN	972	560,486	Random split sample (90:10)	R (caret, xgboost, Keras, pROC)
Kim D et al. [41]	2018	Korea	Prehospital	Critical illness	LR, RF, DNN	5	460,865	10-fold cross-validation	MATLAB, Python (tensorflow)
Klug et al. [42]	2019	Israel	ED	Early and short-term mortality	GBDT	11	799,522	Training (year 2012 to 2017) and Validation (year 2018)	Python, R (XGBoost library)

Kwon et al. [43]	2019	Korea	ED	Critical illness, hospitalization	DNN, RF, LR	8	2,937,078	10-fold cross-validation	Python (TensorFlow), R (glmulti, randomForest)
Levin et al. [44]	2018	USA	ED	Critical care, Hospitalization, emergency procedure	RF	18	172,726	Random split sample (66:33), bootstrapping	
Olivia et al. [45]	2018	India	ED	Triage Level	DT, SVM, NN, NB	8	-	10-fold cross-validation	Python (Keras)
Raita et al. [46]	2019	USA	ED	Critical illness, hospitalization	LR, LASSO, RF,GBDT, DNN	6	135,470	Random split sample (70:30) + 10-fold cross-validation	R (glmnet, ranger, caret, xgboost, Keras)
Rendell et al. [47]	2019	Australia	ED	Hospitalization	B, DT, LR, NN,NB, KNN	11	1,721,294	10-fold cross-validation	START, Python (scikit-learn)
Roquette et al. [10]	2020	Brazil	Pediatric ED	Hospitalization	SVM, ElasticNet, DNN, Catboost, XGBoost	62	499,853	Training (Jan 2015 to Apr 2018) + CV Test (May to Aug 2018)	R version 3.5 and Python version 3.6 (Keras version 2.2.4 with Tensorflow back-end version 1.10.0 for the deep learning part)
Sterling et al. [23]	2019	USA	ED	Hospitalization	NLP (BOW, PV, TM) & NN	1	260,842	Random split sample (50:50)	Deep Learning for Java (DL4J)
van Rein et al. [48]	2019	Netherlands	Prehospital	Critical illness	LR	48	6,859	Separate external validation	R
Wolff et al. [49]	2019	Chile and Spain	Pediatric	Hospitalization	DL, RF, NB, SVM	7	189,718	Hold-out scheme (80:20) + 5-fold cross-validation	-

Chen et al. [50]	2020	Taiwan	All non-trauma adult ED visits	Hospitalization	LR, DNN	15	18,308	10-fold cross-validation	Python Tensorflow and scikit-learn
Ivanov et al. [51]	2021	Ukraine and Canada	ED	ESI triage acuity	XGBoost	-	166,175	-	Java 8.0 (Oracle Corporation) OpenNLP Java library Python Sklearn and SciPy
Tahayori et al. [26]	2020	Australia	ED	Patient disposition	DNN	1	399,876	80%/20% split	TensorFlow Keras Google BERT

Table 1: Characteristics of the selected articles.



For each article, the included variables are shown by a green diamond. Variables that were not included (or not available) and variables for which no clear information was found are shown with red and gray diamonds, respectively.

Figure 3: Predictors/candidate variables included in the selected articles.

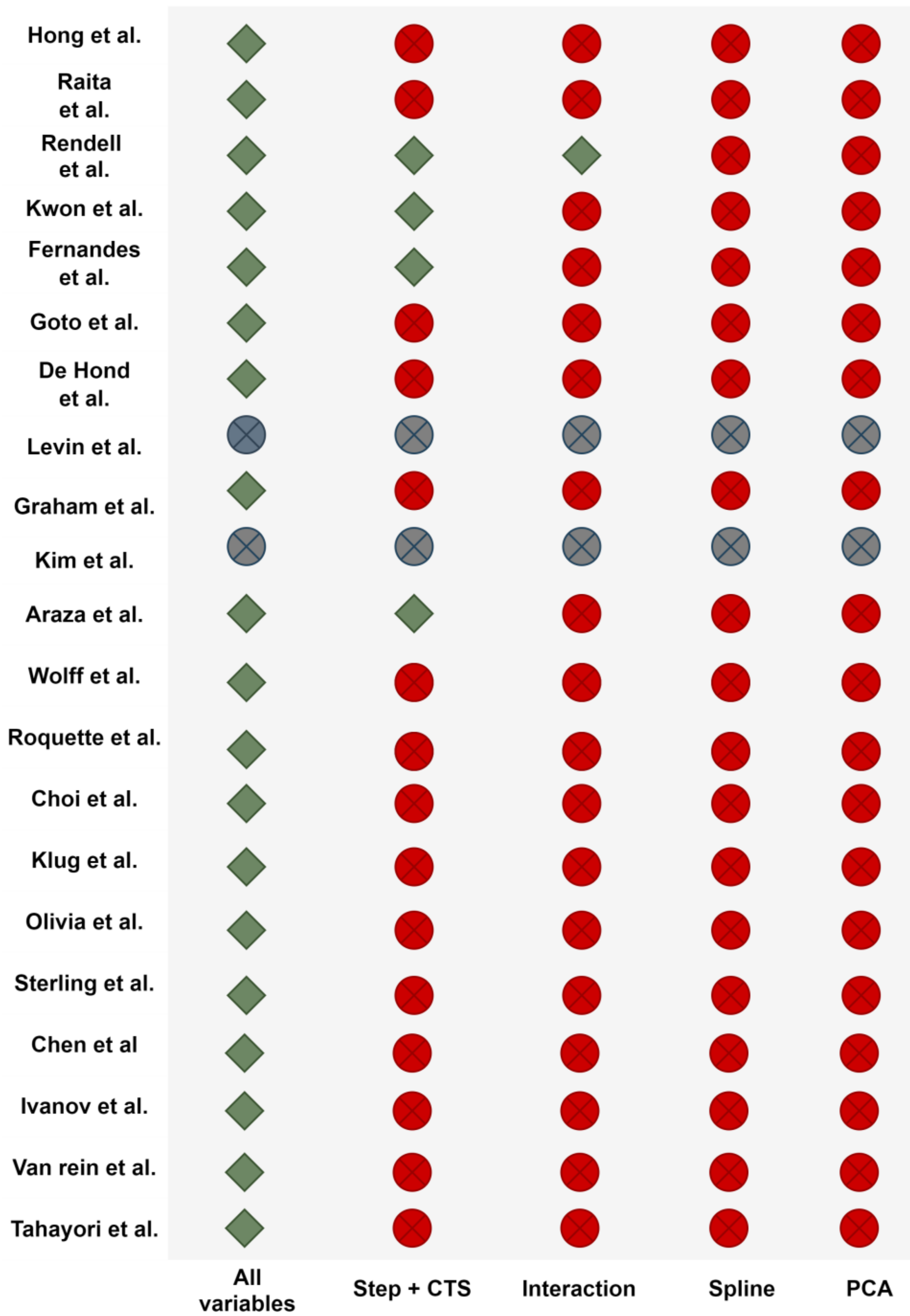


Figure 4: Candidate variable handling and feature engineering for model building in the different studies. Green diamond, yes; red circle with a cross, no; gray circle with a cross, no clear information.

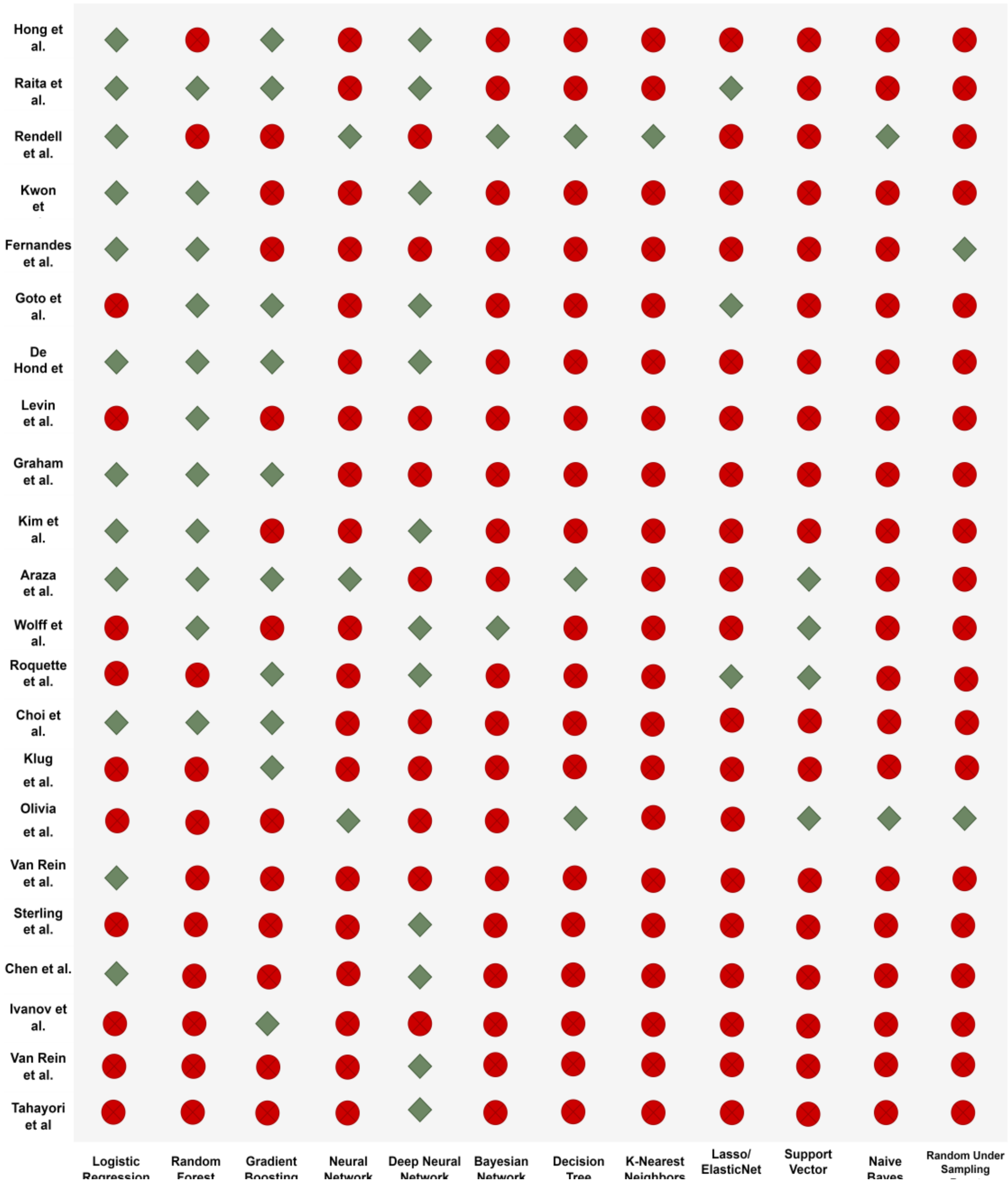
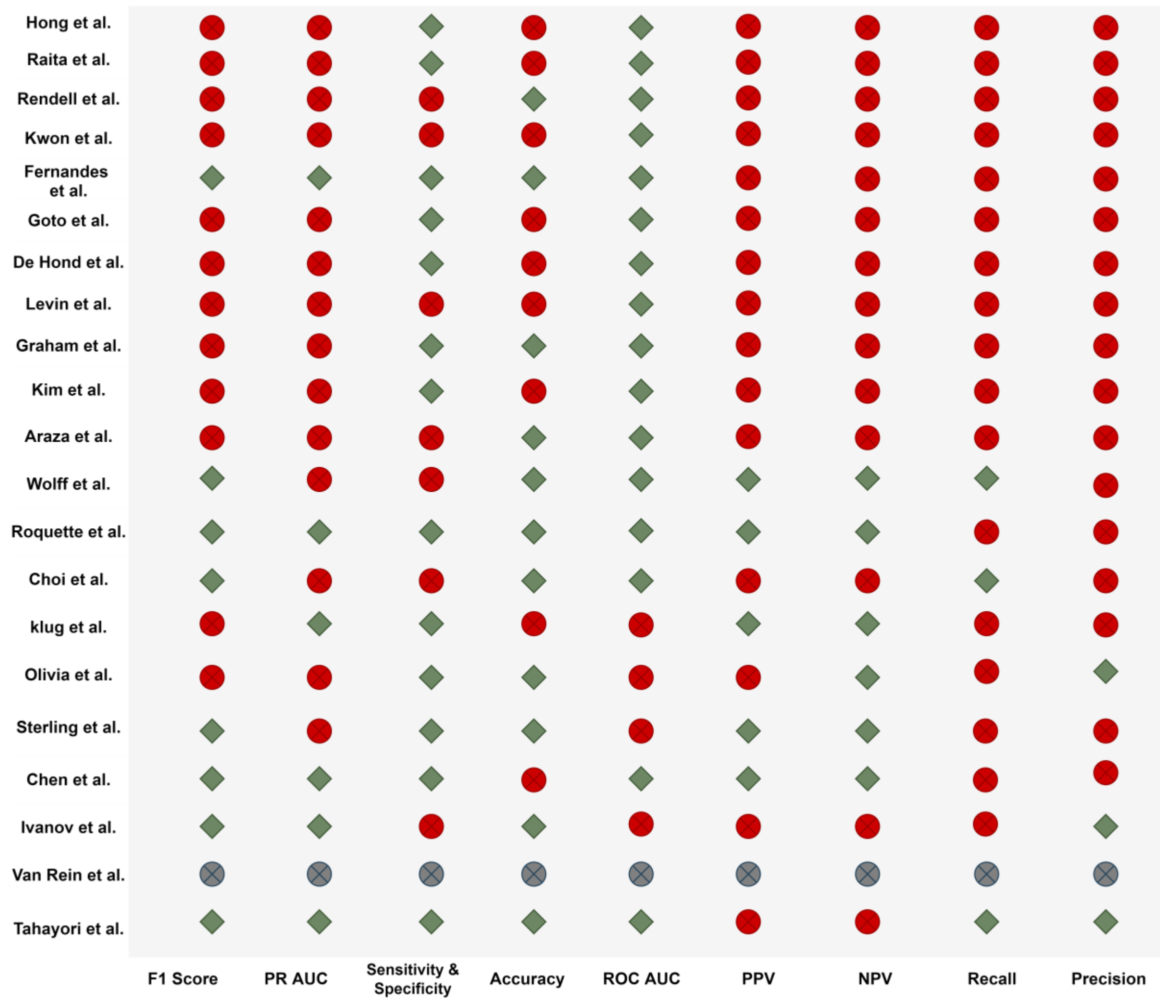


Figure 5: Algorithms used in the selected studies. Green diamond, algorithm used in that study; red circle, not used in that study.



Abbreviations: PR AUC area under the precision recall curve, ROC AUC area under the receiver operating characteristics curve, PPV positive predictive value, NPV negative predictive value.

Figure 6: Evaluation metrics used in the included studies.

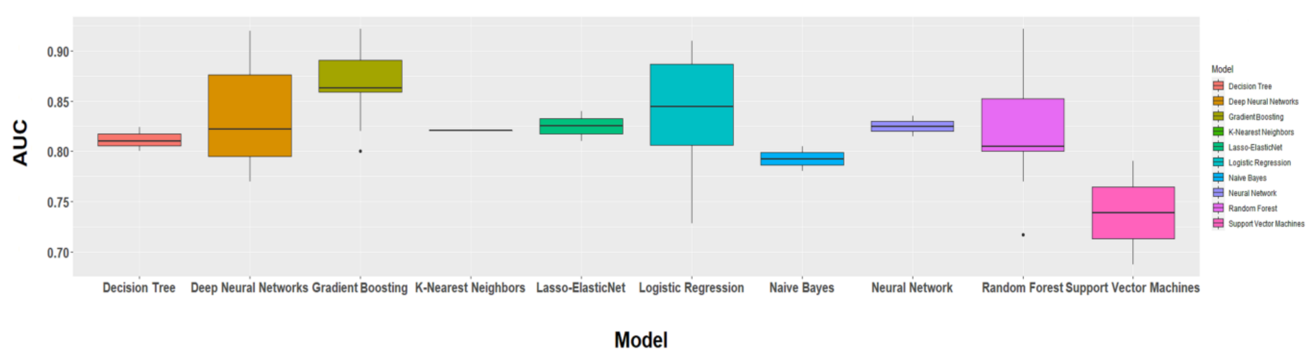


Figure 7: C-statistics of the algorithms used to predict hospitalization.

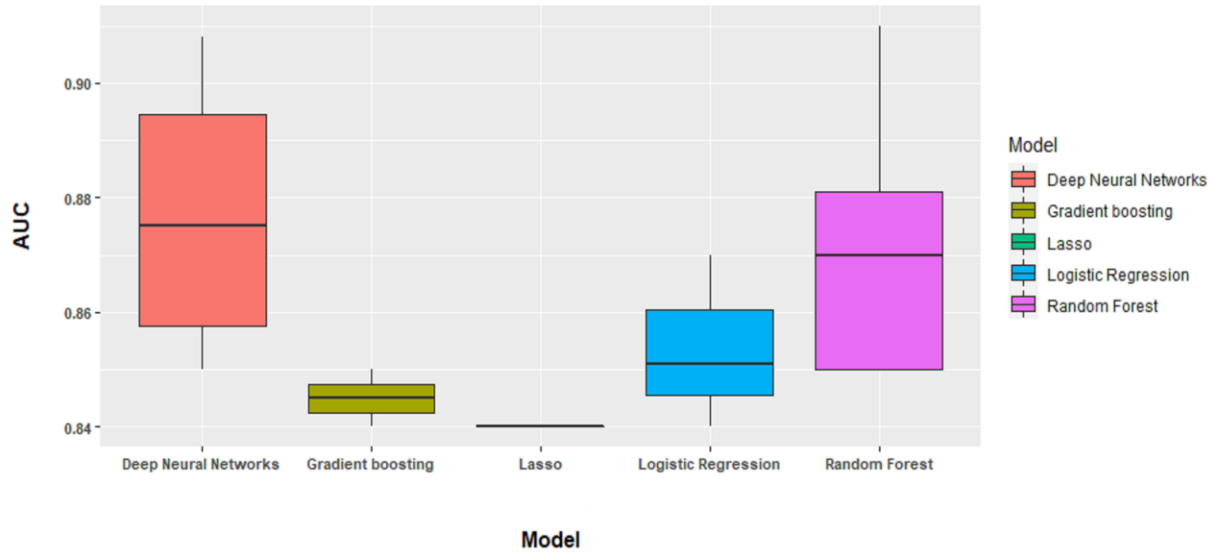
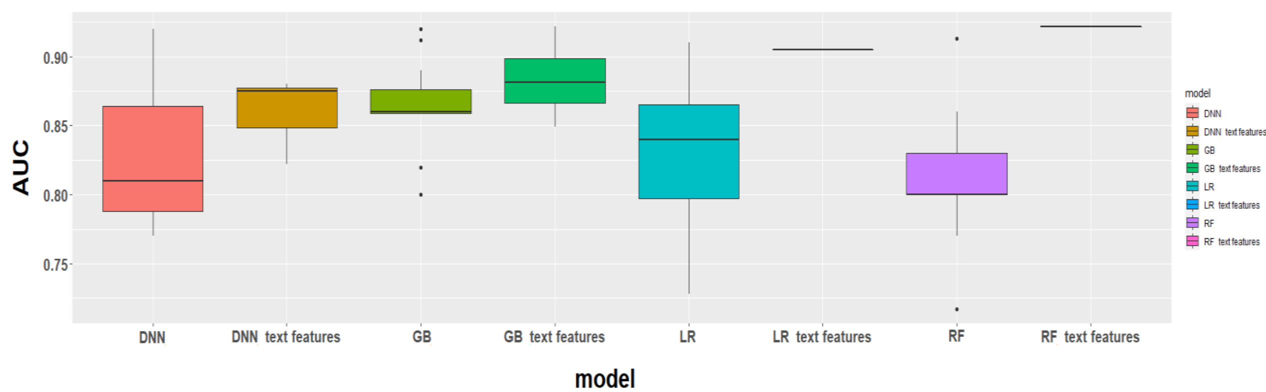


Figure 8: C-statistics of the algorithms used to predict critical illness.



Abbreviations: DNN Deep Neural Networks, GB Gradient Boosting, LR Logistic Regression, RF Random Forest

Figure 9: C-statistics of the algorithms with and without NLP techniques.

4. Discussion

The objective of these studies that developed ED triage algorithms was to propose decision support systems to help health professionals to prioritize high-risk patients. As mentioned in previous review articles [16, 36, 38-40, 41, 43-44, 46], the reference standard on which ED triage is currently based, such as the ESI, can hardly recognize critically ill patients. Indeed, it is hard to deal with such detailed data on the little time available. Advanced AI models based on big volumes of historical clinical data may allow overcoming this obstacle. The aim of the present review was to identify the tools needed to build robust and efficient prediction algorithms that offer higher discrimination performance than the reference standard models. The twenty-one recent and most cited studies from 2018 to

October 2021, selected for this review, described ML-based decision support systems to improve patient triage in ED. Two outcomes were selected: hospital admission and critical illness. The most common methods were Gradient Boosting (17 models), followed by Random Forest, Logistic Regression and Deep Neural Networks (15 models/each).

The objective of this review was not only to describe the developed methods and techniques, but also to identify possible improvements. A common problem with the selected studies was that they did not describe in detail or did not report their feature engineering process. Only one study mentioned that they took into account the predictor interactions [47]. No study explained how they would model non-linear numerical predictors and non-linear relationships (e.g. polynomials or splines). Furthermore, nine of the included studies mentioned that they took into account the hyperparameter calibration [16, 36-40, 46]. However, the majority did not explain the rationale behind the choice of calibration method and did not include the results of this analysis. Yet, the calibration result analysis might be crucial during the development of a transportable model that needs to be adapted to new settings [15, 56-58]. Many authors mentioned the necessity to offer the widest possible range of prediction approaches. For example, Rendell et al. highlighted the different advantages of each ML algorithm and emphasized that these algorithms overcome the limitations of more traditional regression techniques by offering both linear and non-linear decision forms. However, in our selected studies, only two studies implemented six models [10, 36], and most proposed only three to four prediction algorithms.

In all selected studies (n=21), the ED triage outcome was imbalanced in terms of the two disposition classes (hospital admission or not, and critical illness or not). However, only two articles (Wolff et al. [49] and Tarayori et al. [26]) balanced the dataset by over-sampling the minority class to avoid naive results with ‘high accuracy’. Future works could include systematically this step in the ML-based ED triage workflow. The studies on NLP contribution to ED triage performance showed that the addition of textual data enhanced the prediction performance of all studied models, even when using minimal text preprocessing [10, 13, 50]. This suggests that some information included in unstructured data is not captured by the conventional structured triage predictors, even with the minimal text preprocessing. The interest of more recent NLP techniques, such as recurrent neural networks and convolutional neural networks, for textual data could be explored in future studies [13].

Lastly, model-agnostic interpretation methods help to understand how features can affect the model prediction. They are flexible and can be applied to any ML model to find new patterns and to know more about the dataset [59-61]. In the selected studies, the authors used exclusively the Permutation Feature Importance method to identify relevant features. Other methods, such as Partial Dependence Plot, Accumulated Local Effect Plots, Feature interaction (H-statistic), Functional Decomposition, and Global Surrogate Models, could be investigated in future works to identify predictors that might affect the patient triage prediction [59]. Predictive modeling can be useful for human triage as a support tool because expertise in ED triage requires significant training and may be limited in resource-constrained contexts. A support tool based on ML and NLP may be more robust to accurately assess and classify severe symptoms, with the aim of identifying the patients who need rapid treatment in ED.

5. Conclusion

This review found that combining ML with historical clinical data for patient triage in ED has a clear advantage over the reference standard currently in use. However, there is still scope for improvement to enhance the prediction performance and explicability of ML models: 1) integration of predictors' interactions and non-linear relationships; 2) precise information on hyperparameter calibration to make models more transportable; 3) correction of unbalanced datasets using the oversampling method; 4) testing more advanced NLP techniques to convert free text into numerical representation, and 5) more studies on the different model-agnostic interpretation methods to identify predictors that affect the triage process. The goal is to optimize the patient flow in order to improve their management, reduce waiting time, and efficiently use resources [62-63].

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data generated or analyzed during this study are included in this published article. If readers need supplementary information, they can contact me (fei.gao@ehesp.fr).

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

FG designed the project, performed the statistical analysis and drafted the manuscript. SD supervised the overall project, oversaw the statistical analysis, and helped to draft and revised the manuscript. CL, KK and MG performed the statistical analysis with FG and SD. BB gave important suggestions to this study. All authors interpreted the data and reviewed the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript.

Acknowledgements

This research is supported by EHESP Rennes, Univ Rennes, EHESP, CNRS, Inserm, Arènes - UMR 6051, RSMS - U 1309, ECAMO, Hoptial Bichât - APHP and ENSAI. Points of view or opinions in this article are those of the authors and do not necessarily represent the official position or policies of the EHESP Rennes, UMR 6051, RSMS - U 1309, ECAMO, Hoptial Bichât - APHP and ENSAI.

References

1. National Center for Health Statistics (US). Health, United States, 2012: With Special Feature on Emergency Care. National Center for Health Statistics (US) (2013).
2. Rui P, Kang K, Ashman JJ. National Hospital Ambulatory Medical Care Survey: 2016 emergency department summary tables. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention (2016).
3. Greenwald PW, Estevez RM, Clark S, et al. The ED as the primary source of hospital admission for older (but not younger) adults. Am J Emerg Med 34 (2016): 943-947

4. Oberlin M, Andrès E, Behr M, et al. La saturation de la structure des urgences et le rôle de l'organisation hospitalière : réflexions sur les causes et les solutions [Emergency overcrowding and hospital organization: Causes and solutions]. Rev Med Interne 41 (2020): 693-699.
5. Shafaf N, Malek H. Applications of Machine Learning Approaches in Emergency Medicine; a Review Article. Arch Acad Emerg Med 7 (2019): 34.
6. Gottlieb M, Farcy DA, Moreno LA, et al. Triage Nurse-Ordered Testing in the Emergency Department Setting: A Review of the Literature for the Clinician. J Emerg Med 60 (2021): 570-575.
7. Nevill A, Kuhn L, Thompson J, et al. The influence of nurse allocated triage category on the care of patients with sepsis in the emergency department: A retrospective review. Australas Emerg Care 24 (2021): 121-126.
8. Farrohknia N, Castrén M, Ehrenberg A, et al. Emergency department triage scales and their components: a systematic review of the scientific evidence. Scand J Trauma Resusc Emerg Med 19 (2011): 42.
9. Gilboy N, Tanabe P, Travers D, et al. Emergency severity index (esi): a triage tool for emergency department care, version 4. Implementation handbook 2012 Edition (pp. 12–0014). Agency for Healthcare Research and Quality (2011).
10. Roquette BP, Nagano H, Marujo EC, et al. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. Neural Netw 126 (2020): 170-177.
11. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas 30 (2018): 870-874.
12. Blomberg SN, Christensen HC, Lippert F, et al. Effect of Machine Learning on Dispatcher Recognition of Out-of-Hospital Cardiac Arrest During Calls to Emergency Medical Services: A Randomized Clinical Trial. JAMA Netw Open 4 (2021): e2032320.
13. Choi SW, Ko T, Hong KJ, et al. Machine Learning-Based Prediction of Korean Triage and Acuity Scale Level in Emergency Department Patients. Healthc Inform Res 25 (2019): 305-312.
14. Salman OH, Taha Z, Alsabah MQ, et al. A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work. Comput Methods Programs Biomed 209 (2021): 106357.
15. Miles J, Turner J, Jacques R, et al. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. Diagn Progn Res 4 (2020): 16.
16. Goto T, Camargo CA Jr, Faridi MK, et al. Machine learning– based prediction of clinical outcomes for children during emergency department triage. JAMA Netw. Open 2 (2019): e186937.
17. Xie F, Ong MEH, Liew JNMH, et al. Score for Emergency Risk Prediction (SERP): An

Interpretable Machine Learning AutoScore-Derived Triage Tool for Predicting Mortality after Emergency Admissions. medRxiv (2021).

18. Xie F, Liu N, Yan L, et al. Development and validation of an interpretable machine learning scoring tool for estimating time to emergency readmissions. EClinicalMedicine 45 (2022): 101315.
19. Klang E, Kummer BR, Dangayach NS, et al. Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach. Sci Rep 11 (2021): 1381.
20. Hunter-Zinck HS, Peck JS, Strout TD, et al. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. J Am Med Inform Assoc 26 (2019): 1427-1436.
21. Ong ME, Lee Ng CH, Goh K, et al. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. Crit Care 16 (2012): R108.
22. Choi DH, Hong KJ, Park JH, et al. Prediction of bacteremia at the emergency department during triage and disposition stages using machine learning models. Am J Emerg Med 53 (2022): 86-93.
23. Sterling NW, Patzer RE, Di M, et al. Prediction of emergency department patient disposition based on natural language processing of triage notes. Int. J. Med. Inform 129 (2019): 184-188.
24. Zhang X, Kim J, Patzer RE, et al. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf. Med 56 (2017): 377-389.
25. Dinh MM, Russell SB, Bein KJ et al. The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. BMC Emerg. Med 16 (2016): 46.
26. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. Emerg Med Australas (2020).
27. Moher D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine 6 (2009): e1000097.
28. Pombo N, Araújo P, Viana J. Knowledge discovery in clinical decision support systems for pain management: a systematic review. Artif Intell Med 60 (2014): 1-11.
29. Fernandes M, Vieira SM, Leite F, et al. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review. Artif Intell Med 102 (2020): 101762.
30. Pereira CR, Pereira DR, Weber SA, et al. A survey on computer-assisted parkinson's disease diagnosis. Artificial intelligence in medicine (2018).
31. Haddaway NR, Collins AM, Coughlin D, et al. The role of google scholar in evidence reviews and its applicability to grey literature

searching. PloS one 10 (2015): e0138237.
32. Lidal IB, Holte HH, Vist GE. Triage systems for pre-hospital emergency medical services-a systematic review. Scandinavian journal of trauma, resuscitation and emergency medicine 21 (2013): 28.
33. Belard A, Buchman T, Forsberg J, et al. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. Journal of clinical monitoring and computing 31 (2017): 261-271.
34. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 11 (2014): e1001744.
35. Wolff RF, Moons KGM, Riley RD, et al. PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 170 (2019): 51-58.
36. Araz O M, Olsona D, Ramirez-Nafarrateb A. Predictive analytics for hospital admissions from the emergency department using triage information. International Journal of Production Economics (2019): 199-207.
37. De Hond A, Raven W, Schinkelshoek L, et al. Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope?. International Journal of Medical Informatics (2021).
38. Fernandes M, Mendes R, VieiraI S M, et al. Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing. PLOS (2020).
39. Graham B, Bond R, Quinn M, et al. Using Data Mining to Predict Hospital Admissions From the Emergency Department. IEEE Access 6 (2018): 10458-10469.
40. Hong W S, Haimovich A D, Taylor R A. Predicting hospital admission at emergency department triage using machine learning. PLOS (2018).
41. Kim D, You S, So S, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. PLOS (2018).
42. Klug M, Barash Y, Bechler S, et al. A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. Journal of General Internal Medicine (2019): 220-227.
43. Kwon J, Jeon K-H, Lee M, et al. Deep Learning Algorithm to Predict Need for Critical Care in Pediatric Emergency Departments. Pediatric Emergency Care (2019).
44. Levin S, Toerper M, Hamrock E, 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 71 (2018): 565-574.e2.
45. Olivia D, Nayak A, Balachandra M. Machine learning based electronic triage for emergency department. Springer (2021).
46. Raita Y, Goto T, Faridi M K, et al. Emergency department triage prediction of

clinical outcomes using machine learning models. Critical Care. Critical Care (2019).

47. Rendell K, Koprinska I, Kyme A, et al. The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making. Emergency Medicine Australasia (2018).
48. van Rein E A J, van der Sluijs R, Voskens F J, et al. Development and Validation of a Prediction Model for Prehospital Triage of Trauma Patients. Jama Surgery (2019).
49. Wolff P, Ríos S A, Graña M. Setting up standards: A methodological proposal for pediatric Triage machine learning model construction based on clinical outcomes. Expert Systems with Applications (2021).
50. Chen CH, Hsieh JG, Cheng SL, et al. Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. Int J Med Inform 139 (2020): 104146.
51. Ivanov O, Wolf L, Brecher D, et al. Improving ED Emergency Severity Index Acuity Assignment Using Machine Learning and Clinical Natural Language Processing. J Emerg Nurs 47 (2021): 265-278.e7.
52. Jabbar H, Khan RZ. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Science (2015).
53. Lei J. Cross-validation with confidence. Journal of the American Statistical Association.. 115 (2020) :1978-1997.
54. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality, Proceedings of the 26th International Conference on Neural Information Processing Systems, NIPS (2013).
55. Le Q, Mikolov T. Distributed representations of sentences and documents, Proceedings of the 31st International Conference on Machine Learning, ICML (2014).
56. Jon J Williams, Luciana S Esteves. Guidance on Setup, Calibration, and Validation of Hydrodynamic, Wave, and Sediment Models for Shelf Seas and Estuaries. Advances in Civil Engineering (2017).
57. Tomasz Konopka. Correcting machine learning models using calibrated ensembles with ‘mlensemble’. bioRxiv (2021).
58. Fehr J, Piccininni M, Kurth T, et al. A causal framework for assessing the transportability of clinical prediction models. medRxiv (2022).
59. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.) (2022).
60. Liu X, Taylor MP, Aelion CM, et al. Novel Application of Machine Learning Algorithms and Model-Agnostic Methods to Identify Factors Influencing Childhood Blood Lead Levels. Environ Sci Technol 55 (2021): 13387-13399.
61. Neves I, Folgado D, Santos S, et al. Interpretable heartbeat classification using local model-agnostic explanations on ECGs. Comput Biol Med 133 (2021): 104393.
62. Barnes S, Hamrock E, Toerper M, et al. Real-time prediction of inpatient length of stay for discharge prioritization. Journal of the American Medical Informatics Association (2015): e2-e10.

63. Fairley M, Scheinker D, Brandeau ML. Improving the efficiency of the operating room environment with an optimization and machine learning model. Health Care Manag Sci 22 (2019): 756-767.