Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240560

Experimentation of AI Models Towards the Prediction of Medium-Risk Emergency Department Cases Disposition Outcome

Styliana Siakopoulou[a], Antonis Billis[a1], Evangelos Logaras[a], Veroniki Stelmach[b], Maria Zouka[b], Varvara Fyntanidou[b] and Panagiotis Bamidis[a]
[a] *Lab of Medical Physics and Digital Innovation, School of Medicine, Aristotle University of Thessaloniki, Greece*
[b] *AHEPA General University Hospital of Thessaloniki, Greece*
ORCiD ID: Styliana Siakopoulou https://orcid.org/0009-0002-8528-8017

Abstract. The overwhelming volume of patients in emergency departments (EDs) is a significant problem that hinders the delivery of high quality healthcare. Despite their great value, triage protocols are challenging to put into practice. This paper examines the utility of prediction models as a tool for clinical decision support, with a focus on medium-severity patients as defined by the ESI algorithm. 689 cases of medium-risk patients were gathered from the AHEPA hospital, evaluated, and their data fed three classifiers: XGBoost (XGB), Random Forest (RF) and Logistic Regression (LR), with the prediction goal being the outcome of their visit, i.e. admission or discharge. Essential features for the prediction task were determined using feature importance and distribution analysis. Despite having many missing values or high sparsity datasets, several symptoms and metrics are recommended as crucial for outcome prediction. When fed the patients' vital signs, XGB achieved an accuracy score of 91.30%. Several chief complaints were also proven beneficial. Prediction models can, in general, not only lessen the drawbacks of triage implementation, but also enhance its delivery.

Keywords. Emergency Department, XGBoost, Emergency Severity Index, Clinical Decision Support.

1. Introduction

Overcrowding in the ED is a global phenomenon that negatively impacts the delivery of healthcare services in every way. Poorer clinical decision-making over discharge planning and admission, higher rates of morbidity, mortality and revisits, overall worse care quality, and unwise use of resources are all associated with it [1].

A core clinical task at the ED context is the patients' triage. According to the Emergency Severity Index (ESI), patients are ranked on a 5-level scale; level 1 denotes resuscitation and level 5 non-urgent care, considering the patient's vital signs, demographics, medical history, and chief complaints [2]. The triage protocol is not always effectively applied, though. Its implementation in the EDs is influenced by several factors, including limited access to crucial supporting information, a lack of

[1] Corresponding Author: Antonis Billis; E-mail: ampillis@auth.gr.

clinical experience, shortage of resources or personnel, a heavy workload, no motivation to put in place the system due to inconsistent use, etc. ESI-level 3 cases displayed the greatest challenges since their traits are typically more heterogenous, thus they are harder to prioritize. It would be ideal to review these cases to ensure they have a fair priority, but given the hurdles related to triage this is not always feasible [3].

Computer-aided triage systems can leverage complex clinical data to outrun triage limitations and enhance the overall standard of care. They aid in prioritization in an ED, identifying hidden patterns in massive amounts of data, and offering clinicians unbiased opinions [4,5]. With remarkably positive results in terms of accuracy and validity, machine learning has proven to have a major impact on predicting clinical outcomes like ED cases disposition. This may be extended to distinguish critically sick from stable patients and identify the outcome forecasts for medium-acuity patients to reevaluate their ESI score more easily [6,7].

The purpose of this research is to experiment with various AI-based prediction algorithms that may enhance hospital resource performance and management by increasing the prognosis of medium-level cases' outcomes. Based on events from ED occurrences, focusing on cases that have been triaged and assigned an ESI level 3, the proposed algorithm predicts the disposition outcome of an ED case. This study is a continuation of previous research [4], extending our experimentation to a dataset retrospectively collected within a Greek tertiary hospital environment.

In what follows: Section 2 explains the methodology and Section 3 presents key experimental results. Section 4 discusses the results and Section 5 concludes this work.

2. Methods

2.1. Initial ED dataset

The dataset used in this study was collected during several ED shifts from the AHEPA hospital, including medium risk cases as triaged upon arrival at the ED. Health professionals with minimum of ten years of experience reviewed the cases and certified that the ESI score assigned to the cases in our dataset was accurate. For each case, all information like age, vital signs, pain level, the number of lab tests performed within a year, symptoms upon arrival, and case disposition outcome were collected via a questionnaire developed with inputs from ED experts and based on a public dataset delivered in [8]. A total of 689 cases that were given an ESI score of 3 were included in this study, consenting on sharing their data upon arrival at the ED (AUTH Bioethics Committee, protocol approval no 6222/29-07-2020).

To tackle the challenges of sparsity of the symptoms' dataset, and the rest missing values, two strategies were employed. Regarding symptoms a network of symptoms was developed for feature engineering, and imputation techniques were applied to the rest of the dataset's variables.

2.2. Feature Engineering and data preparation for algorithms' experimentation

Symptoms were mapped into a network that is then visualized as a graph. Each node in the graph represents a case, and each edge connecting two nodes indicates that they share at least one symptom. The weight of each edge shows the total number of symptoms that these cases share. The symptoms, therefore were converted into a bipartite graph, and then its weighted projection was the symptoms' network [9].

The graph's architectural features were calculated using an algorithm for recursively extracting ego-net features (e.g. number of internal and external edges) and node features (e.g. degree). This algorithm has been proven more intuitive than the commonly used communities of nodes [10]. Extracted features included roles and the graph features. Roles contained each node's structural role and the relevant computational information. The graph features included the degree, internal and external edges for each node, as well as the mean of the degree and the internal and external edges for each recursive step in the algorithm.

Approximately 70% of data from the variables (except for symptoms) were missing. We used imputation techniques to cope with the missing data. Multivariate imputers estimate missing data repeatedly, and generally produce more accurate results. We utilized two multiple imputation methods: the iterative imputer that predicts the missing data in a round-robin way, and the KNN imputer, based on the KNN algorithm [11].

2.3. Experimentation

Our aim was predicting the disposition outcome (hospitalization or discharge) based on three distinct sub-datasets formulated from the initial dataset, as follows: i) first, with the symptoms upon arrival, ii) second, with all cases' variables as explained in section 2.1 except for symptoms and iii) third one, that includes the symptoms' network features and the rest variables (again excluding the symptoms). To this end, three algorithms were employed for each one of the three sub-datasets: RF, XGB, LR, implemented in Python. The split ratio of the datasets for training and testing was 80/20 respectively. Validation tests and fine-tuning were performed to optimize the accuracy and enhance the overall performance of the models while avoiding over/under-fitting. Following fine-tuning, the optimal hyperparameter values were chosen and some of the models' default values were replaced. XGB's hyperparameters were gamma (13), eta (0.21), colsample_bytree (0.2). For RF hyperparameters were n_estimators (310), max_features (55), min_samples_split (8), mean_samples_leaf (4). Lastly, for LR, the following hyperparameters were chosen: C (0.7), solver (saga), penalty (l1), class_weight (balanced), fit_intercept (False). Hyperparameters not mentioned were left with default values.

3. Results

3.1. Exploratory Analysis

A distribution analysis of the main variables was performed among the two classes. Out of the 689 cases, 372 (53.99%) were discharged and 317 (46.01%) were hospitalized. Features of discriminant value seemed to be age, oxygen saturation, number of imaging, blood, and biomechanical test within the last year; and symptom network features (degree; external edges). The ten most important symptoms were general weakness, injury complications, generalized abdominal pain, vomiting, headache, vertigo/dizziness, tachypnea/dyspnea, back symptoms, atrial fibrillation, fever. Iron deficiency anemia also scored high and was exclusively present in hospitalized patients (6.56%). Three of the most significant chief complaints/symptoms that also had a high number of occurrences are indicatively displayed in Figure 1.

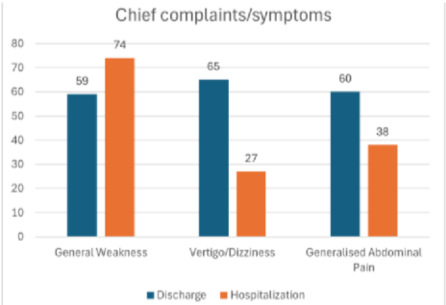

Figure 1. Three most important chief complaints distribution among discharged and admitted ED cases.

3.2. Prediction Outcomes

The classifiers initially ran with their default parameter values, and during validation there was some overfitting and underfitting, which led to fine-tuning. The second sub-dataset (including all variables except for the symptoms) appeared to be the one with the best predictive value when compared to the rest. To maintain some stability between tests, the validation and tuning were first conducted using the same random state of the dataset split. The selection process considered both accuracy and best fit. After choosing the tuned hyperparameter values, more random states were generated to ensure that these values would be optimal in multiple dataset states. Despite its sparsity, the symptoms dataset demonstrated promising results in our prediction task; it provided the best fit for RF and LR, with accuracies of 73.91% and 70.29% respectively. The traits dataset with iterative imputation had the best scores, and the best fit was with the optimized XGB: it produced an accuracy of 91.30%, 0.9 precision, 0.93 recall and 0.91 F1 score (Table 1).

Table 1. Accuracy scores with the selected hyperparameter values

Classifiers	Sub-Datasets Selected	Accuracy Scores
XGB	Traits – Iterative	91.30 %
RF	Symptoms	73.91%
LR	Symptoms	70.29%

4. Discussion

The primary goal of this study was to identify predictive methods and clinical variables that are important in predicting the disposition outcome of an ED case that has been triaged as medium-acuity. Three state-of-the-art classifiers (XGB, RF, and LR) were used for this purpose, and their accuracy ranged roughly from 70% to 92%. Compared to previous work [4], the current models developed with a new dataset produced more promising results. The data were collected from a Greek tertiary hospital and inspected by medical professionals to increase confidentiality, notwithstanding the possibility of human error. This is only a strong hypothesis generation, as the dataset has a lot of missing data and the results are prejudiced. Further, we found more potentially beneficial variables for a dataset intended for this kind of prediction. In the future, the dataset will be enriched by merging the most significant features with the remaining clinical data.

5. Conclusions

EDs are far too overcrowded, managing patients who need triage is not ideal, and there is a chance that medium-risk patients will be wrongly classified. A computer-aided system for these cases is required, as it might enhance the care provided. Though the results of our study are encouraging, more work and data are needed to yield more robust results. Future research will examine the value of more variables in our dataset for the prediction task, feed our data into neural networks, and extract additional features.

Acknowledgements

This work has been partially co-financed by the European Regional Development Fund of the EU and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call Research– Create– Innovate (IntelTriage project, No. T1EDK-02489).

References

[1] Sartini M, Carbone A, Demartini A, Giribone L, Oliva M, Spagnolo AM, et al. Overcrowding in Emergency Department: Causes, consequences, and solutions—A narrative review. Healthcare (Basel) [Internet]. 2022 [cited 2024 Jun 6];10(9):1625. Available from: https://pubmed.ncbi.nlm.nih.gov/36141237/

[2] Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity index triage algorithm version 2 is reliable and valid. Acad Emerg Med [Internet]. 2003 [cited 2024 Jun 6];10(10):1070–80. Available from: https://pubmed.ncbi.nlm.nih.gov/14525740/

[3] Janssen MAP, van Achterberg T, Adriaansen MJM, Kampshoff CS, Schalk DMJ, Mintjes-de Groot J. Factors influencing the implementation of the guideline Triage in emergency departments: a qualitative study. J Clin Nurs [Internet]. 2012 [cited 2024 Jun 6];21(3–4):437–47. Available from: https://pubmed.ncbi.nlm.nih.gov/22171544/

[4] Ketseridou SN, Logaras E, Billis A, Zouka M, Fyntanidou B, Bamidis P. AI-driven prediction for the disposition of medium-risk incidents visiting emergency departments. In: 25th Pan-Hellenic Conference on Informatics. ACM; 2021:490-494. doi:10.1145/3503823.3503912

[5] Billis A, Zouka M, Nicopolitidis P, Lagakis P, Logaras E, Karanasiou N, et al. Towards the definition of an intelligent triage and continuous monitoring system for hospital Emergency Departments and clinics. Stud Health Technol Inform [Internet]. 2019 [cited 2024 Jun 6];264. Available from: https://pubmed.ncbi.nlm.nih.gov/31438270/

[6] Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care [Internet]. 2019 [cited 2024 Jun 6];23(1). Available from: https://pubmed.ncbi.nlm.nih.gov/30795786/

[7] Riordan JP, Dell WL, Patrie JT. Can patient variables measured on arrival to the emergency department predict disposition in medium-acuity patients? J Emerg Med [Internet]. 2017 [cited 2024 Jun 6];52(5):769–79. Available from: https://pubmed.ncbi.nlm.nih.gov/28012828/

[8] Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One [Internet]. 2018;13(7):e0201016. Available from: http://dx.doi.org/10.1371/journal.pone.0201016

[9] Stankova M, Praet S, Martens D, Provost F. Node classification over bipartite graphs through projection. Mach Learn [Internet]. 2021;110(1):37–87. Available from: http://dx.doi.org/10.1007/s10994-020-05898-0

[10] Henderson K, Gallagher B, Eliassi-Rad T, Tong H, Basu S, Akoglu L, et al. RolX: Structural role extraction & mining in large graphs. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM; 2012.

[11] Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: A tutorial on multiple imputation. Can J Cardiol [Internet]. 2021 [cited 2024 Jun 6];37(9):1322–31. Available from: https://pubmed.ncbi.nlm.nih.gov/33276049/