Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220775

Exploratory Clustering for Emergency Department Patients

Georgios FERETZAKIS[a,b,1], Aikaterini SAKAGIANNI[c], Dimitris KALLES[a], Evangelos LOUPELIS[d], Lazaros TZELVES[e], Vasileios PANTERIS[f], Rea CHATZIKYRIAKOU[g], Nikolaos TRAKAS[h], Stavroula KOLOKYTHA[i], Polyxeni BATIANI[i], Zoi RAKOPOULOU[j], Aikaterini TIKA[j], Stavroula PETROPOULOU[c], Ilias DALAINAS[j] and Vasileios KALDIS[i]

[a] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[b] *Department of Quality Control, Research and Continuing Education, Sismanogleio General Hospital, Marousi, Greece*
[c] *IT department, Sismanogleio General Hospital, Marousi, Greece*
[d] *Intensive Care Unit, Sismanogleio General Hospital, Marousi, Greece*
[e] *Second Department of Urology, Sismanogleio General Hospital, Marousi, Greece*
[f] *Gastroenterology Department, Sismanogleio General Hospital, Marousi, Greece*
[g] *Hematology Laboratory, Sismanogleio General Hospital, Marousi, Greece*
[h] *Biochemistry Department, Sismanogleio General Hospital, Marousi, Greece*
[i] *Emergency Department, Sismanogleio General Hospital, Marousi, Greece*
[j] *Administration, Sismanogleio General Hospital, Marousi, Greece*

Abstract. Emergency department (ED) overcrowding is an increasing global problem raising safety concerns for the patients. Elaborating an effective triage system that properly separates patients requiring hospital admission remains difficult. The objective of this study was to compare a clustering-related technique assignment of emergency department patients with the admission output using the k-means algorithm. Incorporating such a model into triage practice could theoretically shorten waiting times and reduce ED overcrowding.

Keywords. Machine learning, unsupervised learning, emergency department, hospital admission, clustering, k-means

1. Introduction

Overcrowding of Emergency Departments is increasingly encountered nowadays, raising safety concerns for the patients. Emergency department physicians are asked to decide whether people visiting the emergency department have a medical problem serious enough to warrant admission to the hospital, or they can be safely discharged. Effective triage systems that differentiate and give precedence to severely ill patients are still lacking.

Recently, several studies describing effective triage prediction in ED using machine learning techniques have been published [1,2].

[1] Corresponding Author, Georgios FERETZAKIS, PhD, School of Science and Technology, Hellenic Open University, Patras, 263 35, Greece; E-mail: georgios.feretzakis@ac.eap.gr

The implementation of data mining and machine learning techniques, such as clustering, in the Emergency Department (ED) may help clinicians detect the patient subgroups which are most suitable for admission, according to disease severity and the need for immediate treatment. This preliminary study uses the k-means algorithm to cluster ED patients into admission or discharge subgroups based on demographic and laboratory parameters. Segmenting the total ED patient flow into distinct clusters by exploiting laboratory and demographic data might be a novel and prototype approach that could support the healthcare system, reduce waiting times, fasten the triage process, and decrease ED congestion.

The objective of this clustering analysis was to present an algorithm with adequate performance to enable healthcare professionals to take rapid and evidence-based decisions during ED triage.

2. Methods and Materials

This observational retrospective study was conducted in the ED of a public tertiary care hospital and has been approved by the Sismanogleio General Hospital's Institutional Review Board. During one year (2020), 13,991 ED visits with routine laboratory testing were recorded.

Our analysis data set includes the following: Creatinine (CREA), C-Reactive Protein (CRP), Creatine Kinase (CPK), Lactate Dehydrogenase (LDH), serum levels of Urea (UREA), D-Dimer, lymphocyte count (LYM%), Activated Partial Thromboplastin Time (aPTT), International Normalized Ratio (INR), hemoglobin (HGB), neutrophil count (NEUT%), white blood cells (WBC) and platelets (PLT), age, and ED outcome (admission or discharge).

For missing data imputation, the Multiple Imputations by Chained Equations (MICE) was used [3,4]. To perform the cluster analysis in R programming environment (version 4.1.1) [5] in RStudio IDE (version 2021.09.0) using the *dplyr*, *ggpubr*, and *factoextra* packages [6-8] and the *k*-means algorithm [9,10]. The *k*-means clustering method takes as input *n* observations and an integer, *k*, indicating the preferred number of clusters into which these *n* observations are partitioned so that each observation belongs to the cluster with the closest mean. The *k*-means algorithm's iterative approach is used to minimize the distance between each observation and the mean of the related cluster.

We applied the R function *FeatureImpCluster()*[11] to improve the clustering quality, which measures feature importance in *k*-means clustering. The importance of a feature is measured by the misclassification rate relative to the baseline cluster assignment due to a random permutation of feature values. The features selected by this process are AGE, CRP, HGB, NEUT%, LYM%, and UREA.

3. Results

The allocation of the 13,991 ED visits into the two clusters, based on *k*-means and relative to eventual hospital admission, is shown in Table 1. The ideal form of a table like the one below would have been one where each row and each column of the cells in italics would contain only one non-zero element. Although we do not see the ideal distribution in our results, which we expected, the formation of the two clusters can be distinguished, and we would say that it is far from the worst-case scenario, which would be a 50-50 split.

Table 1. The allocation of the 13,991 ED visits into the two clusters based on the k-means algorithm relative to the hospital admission

	Clusters		
Admission	1	2	Sum
No	2,249	5,439	7,688
Yes	3,932	2,371	6,303
Sum	6,181	7,810	13,991

The *ggscatter()* function from the *ggpubr* package was used to visualize the clusters alongside the admission output in the following plot (figure 1).

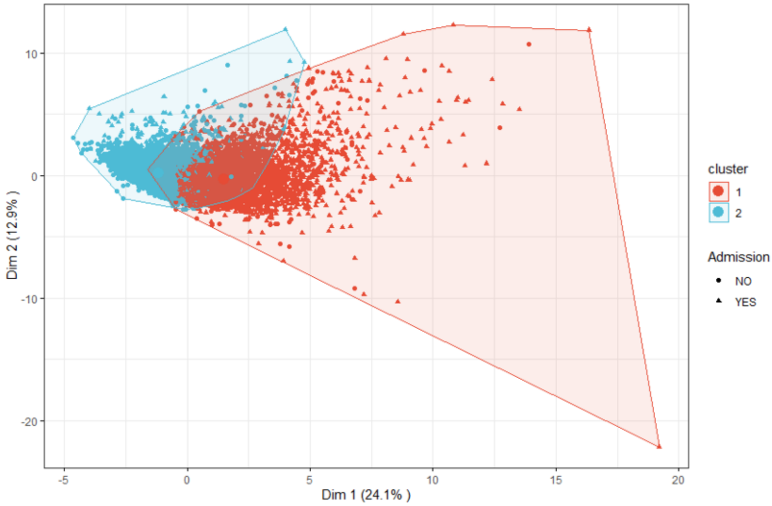

Figure 1. Visualization of the clusters alongside the admission output.

The corresponding R code that was used and the plots produced from this clustering study are available in the online appendix [12].

4. Discussion

Visiting ED is commonly associated with admission to a hospital ward for further management. The physicians of these departments are frequently facing the lack of time, exhausting, long-hour shifts, and mental tiredness that may lead to wrong decisions, namely choosing to admit a mildly ill patient or discharging a critically ill patient. Hardly does this phenomenon has improved in the era of the COVID-19 pandemic since critically ill patients suffering from cardiovascular collapse are constantly attending EDs around the globe.

In an ideal clinical setting, the implementation of algorithms to accurately predict those patients who are critically ill in order to be admitted would certainly ameliorate this problem. In this study, we assessed the usefulness of a complex mathematical approach in order to test whether clustering of patients based on their laboratory values using an algorithm resulted in an accurate prediction.

5. Conclusion

The disagreement between cluster allocation and real-life scenarios is mainly due to the lack of clinical features, such as (for example) functional and mental status impairment indices. Despite the major role of laboratory examinations in patient assessment, the significance of these values for a febrile patient without any underlying comorbidities is certainly different from the one concerning an obese, aged patient suffering from acute decompensation of heart failure. This observation highlights the need for future analysis to take into account additional clinical predictors such as comorbidities and disease severity indices.

References

[1] Feretzakis G, Sakagianni A, Loupelis E, et al. Predicting Hospital Admission for Emergency Department Patients: A Machine Learning Approach. Stud Health Technol Inform. 2022 Jan 14;289:297-300. doi: 10.3233/SHTI210918.
[2] Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019 Feb 22;23(1):64. doi: 10.1186/s13054-019-2351-7.
[3] Van Buuren S, Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 2011;45(3):1–67.
[4] Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv. Methodol. 2001;27:85–96.
[5] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.
[6] Wickham H, Francois R, Henry L, and Muller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.7. 2021.
[7] Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0.
[8] Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.
[9] Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics. 1976;28(1): 100-108. 10.2307/2346830.
[10] Forgy EW. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. Biometrics. 1965(21):768-769.
[11] FeatureImpCluster: Feature Importance for Partitional Clustering. Available at: https://cran.r-project.org/web/packages/FeatureImpCluster/index.html (accessed on 21 March 2022).
[12] Feretzakis G. Exploratory Clustering for Emergency Department Patients. Available at: https://www.learningalgorithm.eu/EDClusterAnalysisICIMTH2022.html (accessed on 28 March 2022).