Predicting Hospital Admission for Emergency Department Patients: A Machine Learning Approach

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

[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] *Thoracic Diseases General Hospital Sotiria, Athens, Greece*
[f] *Second Department of Urology, National and Kapodistrian University of Athens,*
Sismanogleio General Hospital, Marousi, Greece
[g] *Hematology Laboratory, Sismanogleio General Hospital, Marousi, Greece*
[h] *Biochemistry Department, Sismanogleio General Hospital, Marousi, Greece*
[i] *Administration, Sismanogleio General Hospital, Marousi, Greece*
[j] *Emergency Department, Sismanogleio General Hospital, Marousi, Greece*

Abstract. The objective of this study was to establish a machine learning model and to evaluate its predictive capability of admission to the hospital. This observational retrospective study included 3204 emergency department visits to a public tertiary care hospital in Greece from 14 March to 4 May 2019. We investigated biochemical markers and coagulation tests that are routinely checked in patients visiting the Emergency Department (ED) in relation to the ED outcome (admission or discharge). Among the most popular classification techniques of the scikit-learn library through a 10-fold cross-validation approach, a GaussianNB model outperformed other models with respect to the area under the receiver operating characteristic curve.

Keywords. emergency department, machine learning, artificial intelligence, critical care, patient admission, decision support, scikit-learn

1. Introduction and Background

A frequently encountered issue in any Emergency Department (ED) is critical overcrowding that leads to the exhaustion of emergency medical resources. This challenge, still present and exacerbated during the COVID-19 pandemic [1], stems from complex causes. Overcrowding in ED is associated with increased medical errors,

---

increased waiting time, less favorable outcomes, and decreased patient satisfaction [2]. The number of ED visits in the USA increased from 128.97 million to 144.82 million between 2010 and 2016. Expressed as a population rate, ED visits per 1000 persons rose from 416.92 in 2010 to 448.19 in 2016 [3] and continues to increase faster than the population growth [3]. With the introduction of the electronic medical record (EMR) system, ED staff is confronted with a continuously increasing volume of data and information for each patient, including past ED visits and medical notes, previous laboratory results, and diagnostic imaging reports, and this results in overwhelming the capacity to manage and/or make use of it [4].

In the ED context, the medical personnel have to decide by proceeding rapidly in interpreting clinical data to classify patients and predict outcomes. This is fundamental to the emergency department (ED) clinical decisions and has a direct impact on cost, efficiency, and quality of medical care [2].

There do exist significant potential tools for improvement in ED decision-making through the application of Artificial Intelligence [5]. Machine learning (ML) can help predict admission to a hospital from the ED using variables collected as part of routine ED care. ML tools are cost-effective and may be used to help ED personnel make faster and more appropriate disposition decisions, decrease unnecessary testing, and alleviate ED crowding [4-6].

This study aims to present a low-cost ML approach, based on ED data and laboratory exams, that may contribute to the medical decision for patient hospital admission, further improving access to treatment and quality of care.

2. Materials and Methods

This observational retrospective research was conducted in the ED of a public tertiary care hospital in Greece and has been approved by the Institutional Review Board of Sismanogleio General Hospital (Ref. No 15177/2020, 5969/2021). The raw data was retrieved from a standard Laboratory Information System (LIS) and a hospital information system (HIS). After that, the data was processed by a DBMS software by running multiple SQL queries. For the machine learning part, we used scikit-learn [7], an ML library for Python. During the period (14 March – 4 May 2019), 3,204 ED visits were recorded. We investigated biochemical markers and coagulation tests that are routinely checked in patients visiting the ED in relation to the ED outcome (admission or discharge). The data set includes the following variables: serum levels of Urea (UREA), Creatinine (CREA), Lactate Dehydrogenase (LDH), Creatine Kinase (CPK), C-Reactive Protein (CRP), Complete Blood Count with differential, including white blood cells (WBC), neutrophil count (NEUT%), lymphocyte count (LYM%), hemoglobin (HGB), and platelets (PLT), Activated Partial Thromboplastin Time (aPTT), D-Dimer, International Normalized Ratio (INR), age, gender, ambulance use (Ambulance), triage disposition to ED unit and ED outcome (admission or discharge).

Our binary classification problem is considered with two classes of patients who visit the ED, i.e., those patients who are admitted and the rest who are discharged. We evaluated the most popular classification techniques (Logistic Regression, Linear Discriminant Analysis, k-nearest neighbors, Gaussian Naive Bayes, Decision Tree, Random Forest, Bagging classifier, C-Support Vector, XGBoost, LightGBM, Gradient Boosting, Multi-layer Perceptron) of the scikit-learn library through a 10-fold cross-

validation approach, which was applied to avoid overfitting. Missing values were filled in using the *SimpleImputer* method.

One-hot Encoding (OHE) was used for converting categorical features (gender, triage disposition to ED unit, ambulance use, and ED outcome) to numerical features.

Finally, to improve the performance of the final model, we applied a correlation-based feature selection method by evaluating the relationship between each input feature and the target variable (ED outcome). Based on the feature selection process, the numerical variables CPK, aPTT, INR, and PLT were excluded from our analysis since they had the smallest correlation with the ED outcome variable (after OHE).

The performance metrics of the ML techniques that were evaluated in this study were the following:

-Precision (*sklearn.metrics.precision_score*): The precision is the ratio TP / (TP + FP), where TP is the number of true positives and FP is the number of false positives. The precision refers to the classifier's ability not to categorize as positive a sample that is negative [8].

-Recall (*sklearn.metrics.recall_score*): The recall is the ratio TP / (TP + FN) where TP is the number of true positives and FN is the number of false negatives. The recall refers to the classifier's ability to find all the positive samples [8].

-f1 score (*sklearn.metrics.f1_score*): The F1 score can be interpreted as a weighted average of the precision and recall, with the best value being 1 and the worst being 0 [8].

-Accuracy classification score (*sklearn.metrics.accuracy_score*): Classification accuracy is a metric that measures a classification model's performance by dividing the number of correct predictions by the total number of predictions.

-Balanced accuracy score (*sklearn.metrics.balanced_accuracy_score*): The balanced accuracy in binary and multiclass classification problems to deal with imbalanced datasets. It is defined as the average of recall obtained in each class [9].

-ROC AUC (*sklearn.metrics.roc_auc_score*): Compute Area Under the Receiver Operating Characteristic Curve from prediction scores [10].

3. Results

Among the different classifiers that were evaluated through 10-fold cross-validation, a GaussianNB model with *var_smoothing=2e-6* outperformed other models with respect to ROC AUC. The GaussianNB model that implements the Gaussian Naive Bayes algorithm for classification [11] was built and optimized on the open-source Python toolkit scikit-learn. The likelihood of the features is assumed to be Gaussian:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \, exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The parameters σ_y and μ_y are estimated using maximum likelihood, and the parameter *var_smoothing* is a stability calculation that widens the curve to accommodate more samples that are further away from the distribution mean.

The performance metrics of the GaussianNB model are presented in the following table (Table 1).

Table 1. *Performance metrics of the GaussianNB model (10-fold cross-validation)*

	Precision	Recall	f1 score	Accuracy	Balanced accuracy	ROC AUC
GaussianNB	0.739	0.629	0.603	0.701	0.629	0.806

4. Discussion

Disposition decisions in the Emergency Department by triage physicians are often difficult, especially in cases with medium- or low-severity medical problems. The final physician's clinical judgment, sometimes biased, as well as social and safety issues, determines the decision to admit or discharge a patient referred to the ED. Machine learning techniques are increasingly applied in emergency medicine, especially for diagnosis and outcome prediction, supporting medical decision-making and potentially improving patient care [5, 6]. Still, there are some limitations in their use in routine clinical practice [12].

In the present study, we examined the performance of different ML models to predict emergency department disposition based on a multitude of readily available laboratory data. Among the most popular classification techniques of the scikit-learn library through a 10-fold cross-validation approach, a GaussianNB model outperformed other models with respect to area under the receiver operating characteristic curve and showed promising potential in assisting with predicting admission to the hospital.

This study's limitations were a relatively short time period examined, the fact that all patients' data came from a single study center, and the limited explainability of the proposed ML model. In future studies, additional data will be added by including patients from longer periods than the current study; however, the key take-away message is that physicians can be familiarized with ML techniques by presenting to them, retrospectively, the performance statistics of such models and allowing them to reflect on the prospect of harnessing existing data to compare admission rates over time and on the extent to which existing hospital protocols may be subject to human factors.

References

[1]   Houshyar R, Tran-Harding K, Glavis-Bloom J, Nguyentat M, Mongan J, Chahine C, Loehfelm TW, Kohli MD, Zaragoza EJ, Murphy PM, Kampalath R. Effect of shelter-in-place on emergency department radiology volumes during the COVID-19 pandemic. Emerg Radiol. 2020 Dec;27(6):781-784. doi: 10.1007/s10140-020-01797-y.
[2]   Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med. 2013 Jun;61(6):605-611.e6.
[3]   Lane BH, Mallow PJ, Hooker MB, Hooker E. Trends in United States emergency department visits and associated charges from 2010 to 2016. Am J Emerg Med. 2020 Aug;38(8):1576-1581. doi: 10.1016/j.ajem.2019.158423.
[4]   Moulik SK, Kotter N, Fishman EK. Applications of artificial intelligence in the emergency department. Emerg Radiol. 2020 Aug;27(4):355-358. doi: 10.1007/s10140-020-01794-1.
[5]   Shafaf N, Malek H. Applications of Machine Learning Approaches in Emergency Medicine; a Review Article. Arch Acad Emerg Med. 2019 Jun 3;7(1):34.
[6]   Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas. 2018 Dec;30(6):870-874. doi: 10.1111/1742-6723.13145.
[7]   Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
[8]   Powers D. Ailab . Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol. 2011; 2: 2229-3981
[9]   Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. 2010 20th International Conference on Pattern Recognition. doi: 10.1109/icpr.2010.764
[10] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.
[11] Gaussian Naive Bayes [cited 2021Aug20]. [Internet]. Available from: https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes
[12] Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med. 2019 Aug 16;2:77. doi: 10.1038/s41746-019-0155-4..