Telehealth Ecosystems in Practice
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI230747

Identifying Preventable Emergency Admissions in Hospitals Using Machine Learning

Sarah A. ALKHODAIR[a,1], Norah ALTWAIJRI [a] and Ahmed I. ALBARRAK[b]
[a] *IT Department, CCIS, King Saud University, Riyadh, Saudi Arabia*
[b] *Medical Informatics and E-Learning Unit, Medical Education Department, RCHIP, College of Medicine, King Saud University, Riyadh, Saudi Arabia*
ORCiD ID: Sarah A. Alkhodair https://orcid.org/0000-0001-8428-3092, Norah Altwaijri https://orcid.org/0000-0002-0119-5933, Ahmed I. Albarrak https://orcid.org/0000-0003-3246-454X

Abstract. Overcrowding in EDs has been viewed globally as a chronic health challenge. It is directly related to the increased use of EDs for non-urgent issues, leading to increased complications, long waiting times, a higher death rate, or delayed intervention of those more acutely ill. This study aims to develop Machine Learning models to differentiate immediate medical needs from unnecessary ED visits. A Decision Tree, Random Forest, AdaBoost, and XGBoost models were built and evaluated on real-life data. XGBoost achieved the best accuracy and F1-score.

Keywords. Preventable Emergency Admissions, Machine Learning, Overcrowding

1. Introduction

Overcrowding in emergency departments (EDs) is one of the most challenging issues worldwide, especially with limited resources. Fortunately, a substantial number of ED visits are considered preventable. The United States has 131.3 million ED visits annually, only 38 million of them are injury-related and 21.7 million result in critical care admissions [1]. Existing literature handles overcrowding in EDs by predicting related issues [2], or preventing it using a 24/7 nurse access line [3] or assessment applications [4]. Limited studies applied Machine Learning (ML) to prevent ED overcrowding [5]. This study aims to develop ML models on real-life data to differentiate patients with immediate medical needs from unnecessary ED visits. To the best of our knowledge, this is the first work to tackle the problem of identifying preventable emergency admissions (PEAs) before they arrive at hospitals, regardless of patient's specific medical problems.

2. Methods

A real-life dataset of 399,874 records and 972 features collected between March 2014 and July 2017 was used [6]. Irrelevant features, features with more than 60% missing

[1] Corresponding Author: Sarah A. Alkhodair, E-mail: salkhudair@ksu.edu.sa.

values, and records with missing target values were removed. Random forest feature selection algorithm was applied to select those most relevant to predict PEAs, which includes age, gender, blood pressure, heart rate, respiratory rate, chronic diseases, and chief complaints. The final dataset of 140,364 records and 65 features was split into 80% training, 20% testing sets. SMOTE technique [7] was applied to the training set to handle the class imbalance before building four ML models: Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting Machine (XGBoost), and Adaptive Boosting (AdaBoost). The testing set was used to evaluate the models' classification performance.

3. Results and Discussion

Table 1 shows the results of evaluating the classification performance of the ML models. XGBoost is the best-performing model, achieving the highest accuracy, recall, and F1-score. This can be attributed to its reduced complexity and scalable training method that avoids overfitting. RF performs closely to XGBoost, as both can effectively handle noisy and complex health data. However, XGBoost introduces various optimization techniques which allow it to handle such data better. AdaBoost is more sensitive to noisy data than XGBoost and RF, resulting in comparatively lower accuracy, recall, and F1-score.

Table 1. Models' classification performance evaluation results

Matric\ Model	XGBoost	AdaBoost	RF	DT
Accuracy	**77.7%**	75.3%	77.1%	70.0%
Precision	85.1%	**86.9%**	85.2%	81.7%
Recall	**82.8%**	76.5%	81.7%	73.3%
F1-score	**83.9%**	81.4%	83.4%	77.5%

4. Conclusions

ML models proved ability to tackle the ED overcrowding problem by building and evaluating four ML models: DT, RF, AdaBoost, and XGBoost. Dimensionality reduction and hyperparameter tuning techniques can be investigated to improve the model's performance in future work.

References

[1] Cairns C, Kang K. National Hospital Ambulatory Medical Care Survey: 2020 emergency department summary tables. DOI: https://dx.doi.org/10.15620/cdc:121911.

[2] Sharafat A, Bayati M. PatientFlowNet: A Deep Learning Approach to Patient Flow Prediction In Emergency Departments. IEEE Access. 2021 03;PP:1-1.

[3] Chaiyachati K, Mahraj K, Mrad C, O'Malley C, Balasta M, Snider C, et al. Using design and innovation principles to reduce avoidable emergency department visits among employees of a large academic medical center. Healthcare. 2021 03;9:100514.

[4] Ada. Health. Powered by Ada; 2016. Available from: https://ada.com/.

[5] Lewis M, Elad G, Beladev M, Maor G, Radinsky K, Hermann D, et al. Comparison of deep learning with traditional models to predict preventable acute care use and spending among heart failure patients. Scientific Reports. 2021 01;11.

[6] Egan M, Murar F, Lawrence J, Burd H. Identifying the predictors of avoidable emergency department attendance after contact with the NHS 111 phone service: analysis of 16.6 million calls to 111 in England in 2015–2017. BMJ Open. 2020 03;10:e032043.

[7] Blagus R, Lusa L. SMOTE for High-Dimensional Class-Imbalanced Data. BMC bioinformatics. 2013.