Machine learning to detect recreational drugs use in cardiac intensive care units

N. El Beze1, J.-G. Dillinger1, K. Hamzi2, A. Trimaille3, C. Bouleti4, E. Gail1, G. Bonnet5, S. Toupin2, J.-C. Dib6, T. Levasseur7, R. Rossanaly Vasram8, V. Aboyans9, E. Thevenet10, P. Henry1, T. Pezel1

1Hospital Lariboisiere, Department of Cardiology, Paris, France
2Hospital Lariboisiere, Department of Machine Learning and Research, Paris, France
3Civil Hospital/ Nouvel Hopital Civil, Department of Cardiovascular Medicine, Strasbourg, France
4University Hospital of Poitiers, Department of Cardiology, Poitiers, France
5Haut-Leveque Hospital - University Hospital Centre, Cardiology Intensive Care Unit and Interventional Cardiology, Pessac, France
6Clinique Medico-Chirurgicale Ambroise Pare, Department of Cardiology, Neuilly sur Seine, France
7Intermunicipal Hospital Centre of Frejus Saint Raphael, Department of Cardiology, Frejus, France
8University hospital of Reunion Felix Guyon, Department of Cardiology, Saint Denis, Réunion
9University Hospital of Limoges, Department of Cardiology, Limoges, France
10University Hospital of Fort de France, Department of Cardiology, Fort de France, Martinique

On behalf of ADDICT-ICCU investigators

Funding Acknowledgements: Type of funding sources: Foundation. Main funding source(s): Fondation Coeur et Recherche

Background: While recreational drugs use is known to be a strong risk factor of acute cardiovascular events, there is no available prediction tool to detect it in patients admitted to intensive cardiac care units (ICCU), and urinary testing is not routinely performed. Machine learning (ML) methods could provide an accurate model for detection of recreational drugs use.

Purpose: To assess the feasibility of recreational drugs use detection in patients admitted to ICCU using a ML model with clinical, biological, ECG, and echocardiographic data.

Methods: From 7 to 22 April 2021, a systematic screening for all traditional recreational drugs (cannabis, opioids, cocaine, amphetamines, 3,4-methylenedioxymethamphetamine [MDMA]) was performed by urinary testing prospectively in all consecutive patients admitted to ICCU in 39 French centres. We compared several ML algorithms to detect recreational drugs use from clinical, biological, ECG and echocardiographic data. The study population was divided in an internal (31 centres, 80% of the population) and an external (8 centres, 20% of the population) cohorts. The internal cohort was divided in a training (70% of the cohort) and a testing (30% of the cohort) dataset. XGBoost algorithm was used for feature selection to avoid overfitting, and gain interpretability. Final performance of the algorithms were evaluated on the external validation cohort and compared against a traditional logistic regression model, using receiver operating characteristics (ROC) and precision-recall (PR) curves and area-under-the curves (AUC).

Results: Of 1,499 consecutive included in the cohort (63±15 years, 70% male), 161 (11%) had a positive test for recreational drugs (cannabis: 9.1%, opioids: 2.1%, cocaine: 1.7%, amphetamines: 0.7%, MDMA: 0.6%, Figure 1). Of the patients with a positive test, only 57% reported using recreational drugs. Out of 165 clinical, biological, ECG, and echocardiographic features, 9 variables were selected as being the most important in detecting drugs use: age, systolic pulmonary artery pressure, body mass index, hemoglobin, NT-pro BNP, temperature, active smoker status, heart rate, and mean blood pressure. The random forest model showed the best performance compared with the other ML models (AUROC=0.82; and PR-AUC=0.62, Figure 2), and with a traditional model using logistic regression (AUROC 0.82 vs. 0.80, and PR-AUC 0.62 vs 0.46; both p<0.001) to detect recreational drugs use. The random forest model also exhibited a good performance for detecting recreational drugs use in the external validation cohort (AUROC=0.76 and PR-AUC=0.44).

Conclusions: In a large ICCU registry, our ML algorithm including clinical, biological, ECG and echocardiographic data exhibited a higher performance to detect recreational drugs use than traditional statistical methods. Knowing the high rate of underreporting in this study, it is crucial to validate the performance of this ML tool in further larger studies.
Model development

Figure 1: ML model using clinical, biological, and TTE data for recreational drugs use detection

Population constituting the study cohort N=1,499

Recreational drugs use: 11% (N=161)

Internal cohort N=1,205 (31 centres)

Validation cohort N=294 (8 centres)

165 available features

XGBoost for feature selection

9 variables for model building:
- Age
- Systolic PA pressure
- Body mass index
- Hemoglobin
- NT-pro BNP
- Temperature
- Active smoker status
- Heart rate
- Mean blood pressure

Model performances

Figure 2: Model Performances for recreational drugs use detection

ROC

Precision-Recall

SVM: 0.75
Logistic regression: 0.80
XGBoost: 0.81
Random forest: 0.82