Machine Learning for Predicting Patients at Risk of Prolonged Opioid Use Following Surgery

Behzad Naderalvojoud, Tina Hernandez-Boussard

Background

Opioids are potent analgesics often used to manage pain, including postoperative pain. However, opioids can be highly addictive, even when prescribed correctly and taken as directed. The balance between pain management and opioid misuse is challenging[1]. To improve patient outcomes following surgery, it is crucial to identify patients at risk for prolonged opioid use prior to prescribing pain management regimens. Many studies have identified a limited number of features predictive of postoperative prolonged opioid use from real-world data, but these are isolated studies using non-standardized data from different sources, which limits their generalizability and reliability across populations[2-4]. In this study, we developed and validated five machine learning (ML) models to predict patients at risk of prolonged opioid use in a diverse, multisite cohort by evaluating not only the performance, but also their discrimination and calibration abilities. We address generalizability limitations by using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to identify postoperative patients at higher risk of prolonged opioid use based on preoperative risk factors using ML approaches.

Methods

The analysis and prediction models were implemented using ATLAS[5] instance, which uses a research data repository of electronic health records (EHRs) mapped to the OMOP CDM. The cohort includes adult patients who underwent surgery during an inpatient visit between 2008 and 2019 with at least one opioid prescription 30 days before or after the surgery. Patients were included if they had at least two visits two years before surgery and two visits 30 days to two years after surgery. If a patient had multiple surgeries, only the first event was included. We also excluded patients who had any other surgery from two months to seven months after the index surgery.

The primary outcome was prolonged opioid use, defined as a new opioid prescription within three to six months after surgery. All features were defined based on CDM concepts. As a result, 58,405 features were extracted from the CDM clinical tables as follows: (1) demographic features including gender, age groups, race, and ethnicity, (2) procedures, (3) diagnoses, (4) prescriptions, (5) measurement, (6) condition era, (7) drug era, (8) procedure count, (9) condition count, (10) drug count, (11) measurement count, (12) condition group, and (13) drug group. We removed infrequent covariates with less than 0.001 frequency from the feature set. Five ML algorithms were used to develop our models: Lasso Logistic Regression (LR), Random Forest (RF), AdaBoost (AB), Decision Tree (DT), and Naive Bayes (NB). The models were developed using the PatientLevelPrediction (PLP) package[6] from the OHDSI community, version 4.0.2. A stratified random sampling approach was used to split the dataset into train (80%) and test (20%) sets. To label patients in the target cohort, a 90-day time-at-risk (TAR) was defined between three and six months after surgery. For the ML models, any opioid drug exposure during this period was considered a positive case. To evaluate models, standard metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used. We assessed model discrimination using receiver operating characteristic (ROC) and precision recall curve (PRC) with their area under curve (AUC) values. The calibration curves were developed to evaluate the model outputs in terms of their ability to generate calibrated probabilities for identifying patients at risk of prolonged opioid use. To select the best hyperparameters and prediction threshold, we used 10-fold cross validation on the train set with a grid search strategy. In this study, we focused on the methodological aspects of the research, including model development and evaluation; however, generalizability and external validation are being conducted in our ongoing network study called PORPOISE[7]. This study received approval from the institute's Institutional Review Board (IRB).

Results

Out of 32,382 opioid users in the target cohort, 5,017 (15.49%) were prolonged opioid users with a TAR of 90 to 180 days. The data was split into 6,476 test cases with 1003 prolonged opioid users and 25,906 training cases with 4,014 prolonged opioid users. Table 1 shows the performance of all ML models across the test set at the best threshold determined by cross validation. The NB model achieves the highest accuracy, specificity, and PPV, yet has the lowest AUC of 0.68. This shows that NB tends towards the majority class and is affected by the class imbalance problem.

Table 1. Performance of ML models across the test set at the best threshold determined by cross validation.

Models	ACC	AUC	AUPRC	SENS.	PPV	SPEC.	NPV
Lasso Logistic Regression	0.580	0.742	0.397	0.792	0.240	0.541	0.934
Random Forest	0.581	**0.743**	**0.399**	**0.796**	0.242	0.542	**0.935**
AdaBoost	0.575	0.730	0.390	0.774	0.235	0.538	0.928
Decision Tree	0.561	0.676	0.313	0.711	0.218	0.533	0.910
Naïve Bayes	**0.794**	0.680	0.270	0.325	**0.332**	**0.880**	0.877

Figures 1 and 2 illustrate the ROC and PRC curves of the three best performing models, LR, RF, and AB. Overall, RF and LR achieved the highest AUC (0.74) and AUPRC (0.40) and provided the best model discrimination. All the reported results are available in the Shiny app[8].

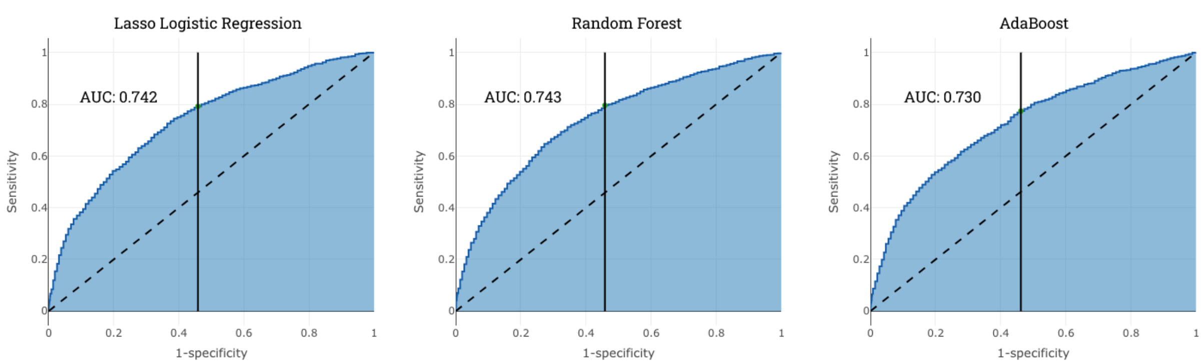

Figure 1. ROC curves for the three ML models on the test set. The black line indicates the performance at the best threshold obtained through cross validation.

[8] https://prolonged-opioid-use-prediction.shinyapps.io/shiny-app/

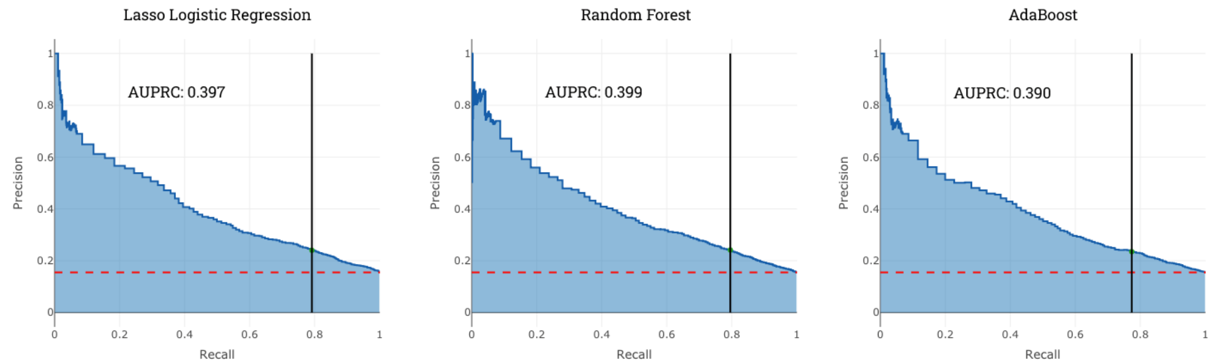

Figure 2. PRC curves for the three ML models on the test set. The black line indicates the performance at the best threshold obtained through cross validation.

Figure 3 illustrates calibration plots for the LR, RF, and AB models over the test set. These plots indicate the accuracy of models in estimating the risk of prolonged opioid use. According to the results, the LR model achieved the best calibration for identifying patients at risk of prolonged opioid use. However, the AB model did not generate calibrated probabilities.

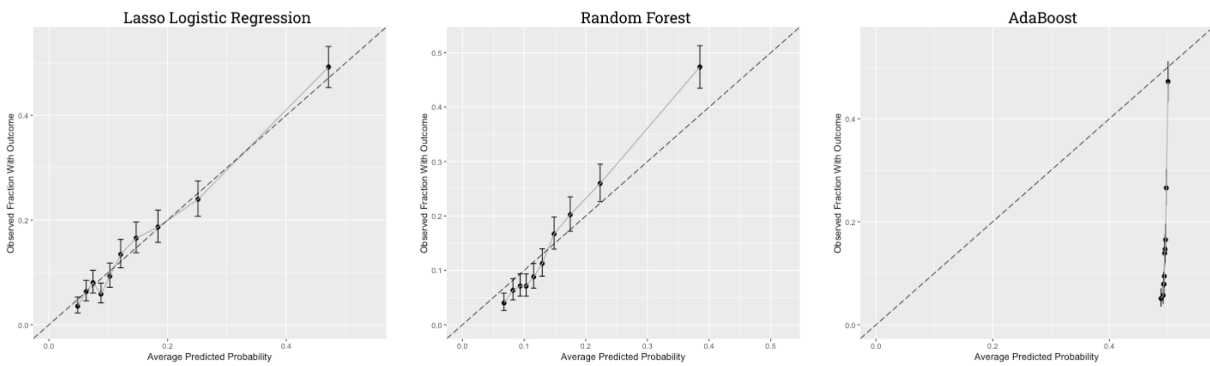

Figure 3. Calibration plots on the test set.

Conclusion

This study proposed ML models to predict patients at risk of prolonged opioid use after surgery using OMOP CDM. LR and RF showed the highest AUC and AUPRC and the most accurately calibrated risk probabilities. However, the LR and RF models had low PPV values, which resulted in increased false positives. On the contrary, NB produced higher PPV and specificity, indicating its tendency toward the majority class. Therefore, incorporating LR (or RF) and NB into a single ensemble model may provide a better and more clinically acceptable balance of sensitivity and specificity when predicting prolonged opioid use after surgery. In addition, the use of calibrated ML models with a wide range of OMOP CDM clinical covariates to predict high-risk patients can lead to more effective postoperative pain management, allowing for early interventions such as reducing opioid dosage or suggesting alternative options. Finally, developing ML models using standardized data and tools improves model transportability to clinical settings as well as reproducibility in training and evaluation across different populations.

Acknowledgment

Research reported in this publication was supported by the National Library Of Medicine of the National Institutes of Health under Award Number R01LM013362. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Brummett CM, Waljee JF, Goesling J, Moser S, Lin P, Englesbe MJ, Bohnert AS, Kheterpal S, Nallamothu BK. New persistent opioid use after minor and major surgical procedures in US adults. JAMA surgery. 2017 Jun 1;152(6): e170504-e170504.
2. Katakam A, Karhade AV, Schwab JH, Chen AF, Bedair HS. Development and validation of machine learning algorithms for postoperative opioid prescriptions after TKA. Journal of orthopaedics. 2020 Nov 1;22:95-9.
3. Dong X, Deng J, Rashidian S, Abell-Hart K, Hou W, Rosenthal RN, Saltz M, Saltz JH, Wang F. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. Journal of the American Medical Informatics Association. 2021 Aug;28(8):1683-93.
4. Ward A, Jani T, De Souza E, Scheinker D, Bambos N, Anderson TA. Prediction of prolonged opioid use after surgery in adolescents: insights from machine learning. Anesthesia & Analgesia. 2021 Jul 14;133(2):304-13.
5. A unified web interface for the OHDSI (observational health data sciences and informatics) tools. URL: https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/
6. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. Journal of the American Medical Informatics Association. 2018 Aug;25(8):969-75. URL: https://github.com/OHDSI/PatientLevelPrediction
7. PORPOISE: Development and external validation of ML models for identifying patients at risk of postoperative prolonged opioid use. URL: https://github.com/ohdsi-studies/PORPOISE