Journal of Pharmaceutical Health Services Research, 2021, Vol 12, 502–508
https://doi.org/10.1093/jphsr/rmab055
Research Paper
Advance Access publication 19 October 2021

JPHSR Journal of Pharmaceutical Health Services Research

OXFORD

Research Paper

Using machine learning to classify patients on opioid use

Shirong Zhao[1], Jamie Browning[2], Yan Cui[3] and Junling Wang[2,*]

[1]Department of Investment, School of Finance, Dongbei University of Finance and Economics, Dalian, Liaoning, China
[2]Department of Clinical Pharmacy and Translational Science, University of Tennessee Health Science Center College of Pharmacy, Memphis, TN, USA
[3]Department of Genetics, Genomics & Informatics, University of Tennessee Health Science Center, Memphis, TN USA

*Correspondence: Junling Wang, Department of Clinical Pharmacy and Translational Science, University of Tennessee Health Science Center College of Pharmacy, 881 Madison Avenue, Room 221, Memphis, TN 38163, USA. Tel: +901-448-3601; Email: jwang26@uthsc.edu

Received February 5, 2021; Accepted October 4, 2021.

Abstract

Objectives High-frequent opioid use tends to increase an individual's risk of opioid use disorder, overdose and death. Thus, it is important to predict an individuals' opioid use frequency to improve opioid prescription utilization outcomes.

Methods Individuals receiving at least one opioid prescription from 2016 to 2018 in the national representative data, Medical Expenditure Panel Survey, were included. This study applied five machine learning (ML) techniques, including support vector machine, random forest, neural network, gradient boosting and XGBoost (extreme gradient boosting), to predict opioid use frequency. This study compared the performance of these ML models with penalized logistic regression. The study outcome was whether an individual lied in the upper 10% of the opioid prescription distribution. Predictors were selected based on Gelberg–Andersen's Behavioral Model of Health Services Utilization. The prediction performance was assessed using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) in the test data. Patient characteristics as predictors for high-frequency use of opioids were ranked by the relative importance in prediction in the test data.

Key findings Random forest and gradient boosting achieved the top values of both AUROC and AUPRC, outperforming logistic regression and three other ML methods. In the best performing model, the random forest, the following characteristics had high predictive power in the frequency of opioid use: age, number of chronic conditions, public insurance and self-perceived health status.

Conclusions The results of this study demonstrate that ML techniques can be a promising and powerful technique in predicting the frequency of opioid use and health outcomes.

Keywords: opioid utilization; machine learning; opioid use frequency; random forest

Introduction

According to the Centers for Disease Control and Prevention (CDC), the recent use of opioid prescriptions in the USA has been continuously declining.[1] From 2012 to 2018, the quantities of opioid prescriptions dispensed in the USA have declined from 255 million to 168 million.[1] The overall national opioid prescribing rate (the average number of prescriptions per 100 persons) has also decreased from 81.3 to 51.4 over the same period.[1] The opioid crisis was officially declared a 'public health emergency' in 2017 by the US President. In the same year, the US Department of Health & Human

© The Author(s) 2021. Published by Oxford University Press on behalf of the Royal Pharmaceutical Society. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

Services developed a 5-point opioid strategy and invested about $900 million to tackle the opioid crisis.[2]

From 2017 to 2018, death rates related to prescription opioids significantly declined by 13.5%.[3] These declines suggest that healthcare providers may have become more cautious when prescribing opioids and that the government policies to combat the opioid epidemic have been effective. However, in 2018, nearly 15 000 persons still died of overdoses involving prescription opioids, which is an average of 41 deaths per day.[3] Moreover, 32% of all opioid overdose deaths are related to prescription opioids.[3] These patterns suggest that there is still a severe national public opioid crisis affecting economic and social welfare in the USA.

Numerous past studies have examined opioid utilization, such as opioid prescription numbers and rates, opioid misuse and daily morphine milligram equivalent dose per prescription.[4–11] In these studies, conventional statistical methods, such as logistic regression and ordinary least squares method, were used to study various opioid utilization topics. However, the logistic regression model's prediction performance may not have been adequately evaluated and such model may not offer the best predictive ability and quality compared with alternative methods such as machine learning (ML).[12] During recent years, with the development of information technology, ML has thrived in predictive analysis due to its advantages, such as capturing non-linear and complicated interactions, minimizing errors between the actual and predicted cases and improving the prediction accuracy. Thus, ML offers an alternative and perhaps a superior approach to building a reliable prediction model.

ML has been widely used in studies addressing many health problems, one of which is opioid utilization.[12–23] Some of these studies compare the prediction performance of ML with conventional methods such as logistic regression. For most studies, they find that ML performs better in prediction than logistic regression.[12, 15, 18, 19, 21] Logistic regression is more widely used to examine the associations between the independent and dependent variables than ML due to its high interpretability of the coefficients; however, logistic regression suffers from a low prediction performance due to the issue of imbalanced data.[24] Imbalanced data refer to the number of observations in some of the outcomes classes that appear much more frequently, resulting in predictive models being biased towards the majority group, but the minority group is usually more important.[25] Although ML methods may also have low predictive power when data are imbalanced, various techniques such as random oversampling, random undersampling and creation of artificial data points, are available to adjust the class distribution of a dataset when using ML methods.[25] These techniques provide improved predictive power of ML compared with logistic regression.

High-frequency opioid use tends to increase an individual's risk of opioid use disorder, overdose and death. When an individual receives too many opioid medications, it is crucial for healthcare providers to identify and offer fewer prescriptions or alternative therapies to reduce the risks of opioid abuse. Thus, it is important to predict an individuals' opioid use frequency to improve opioid utilization patterns. However, no previous study has used ML techniques to predict high-frequency opioid utilization in a national representative sample for the US population. The objectives of this study were: (1) to apply five different ML techniques to predict opioid utilization, (2) to compare the performance of ML with logistic regression to determine if ML can offer additional predictive power and select the best performing ML technique based on performance measures and (3) to evaluate the relative importance of each variable in explaining opioid utilization in the best performing ML technique. The study results may provide insights into predictive power of ML techniques in high-frequency opioid utilization. Additionally, the findings may offer vital information to healthcare providers and policymakers regarding the essential factors affecting opioid utilization.

Methods

Data source

This study analysed data from 3 years (2016–2018) of the Medical Expenditure Panel Survey (MEPS).[26] The MEPS contains national representative data of healthcare use, expenditure, payment sources and more for the non-institutionalized civilian population in the USA. MEPS uses questionnaires to collect information from individual household members. The Full-Year Consolidated Data File, the Medical Conditions File and the Prescribed Medicines File from MEPS were analysed in this study.

Outcome variable

The outcome variable in this study was a binary variable measuring opioid use frequency. A list of opioid drugs compiled by the CDC was used to identify opioid prescriptions in the Prescribed Medicines File of MEPS.[27] For each patient, the number of opioid prescriptions per year was computed. Although opioid use is complex, information available in the MEPS is not conducive to the inclusion of some other outcomes such as dose or morphine milligram equivalents of prescriptions. Further, this study is not meant to be all-inclusive. Individuals who received at least one opioid prescription were included in the study sample. Non-users of opioids were excluded because they may have different characteristics than the users. In total, there were 7915 patients in the study sample. Individuals filled 11 opioid prescriptions at the 90th percentile. The outcome variable was coded '1' for individuals receiving no less than 11 prescriptions and '0' otherwise.

Independent variables

Gelberg–Andersen's Behavioral Model for Vulnerable Populations (hereafter Gelberg–Andersen Model) was utilized to select variables affecting opioid use frequency.[28] According to this model, health service utilization by an individual is affected by three groups of factors: predisposing, enabling and need factors. The variables that affect the propensity to opioid use were considered predisposing factors, including male (yes/no), age, race/ethnicity, married (yes/no) and education. Race/ethnicity included non-Hispanic Whites, non-Hispanic Blacks, Hispanics, non-Hispanic Asians and Others. Education measured whether an individual's education level was higher than a high school degree. The enabling factors included insurance status, income and census region. Insurance status included three categories: any private insurance, public insurance only and no insurance. Income consisted of four categories: poor or near-poor (income <125% of the poverty line), low income (income ≥125% but <200% of the poverty line), middle income (income ≥200% but <400% of the poverty line) and high income (income ≥400% of the poverty line). Census region included the Northeast, Midwest, South and West regions.

The need factors included an individual's self-perceived health status (poor, fair, good, very good and excellent) and the total number of chronic conditions. The total number of chronic conditions was computed as the row count of chronic conditions based on ICD10CDX in MEPS. Starting in 2016, MEPS began using ICD10CDX based on ICD-10-CM to replace ICD9CODX for

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

coding medical conditions. A reference dataset of chronic condition indicators from the Healthcare Cost and Utilization Project was employed to identify chronic conditions in the Medical Conditions of MEPS.[29] The number of chronic conditions was classified into three levels: ≤1, 2–4 and ≥5.

Model development

Five different ML models were applied to predict high-frequency opioid users: support vector machine, random forest, neural network, gradient boosting and XGBoost (extreme gradient boosting). These five models were chosen due to their respective advantages, such as popularity, capturing non-linear associations and resiliency to overfitting. To evaluate the performance of ML models, a fitted penalized logistic regression was used as a comparison. Instead of using the conventional logistic regression (i.e. without shrinking the regression coefficients towards zero), the penalized logistic regression was utilized due to the outcome variable's imbalanced problem.[12] All models were implemented in JupterLab with Python Kernel. The *Scikit-learn* package was utilized for penalized logistic regression, support vector machine, random forest, neural network and gradient boosting while the *xgboost* package was employed for XGBoost.

The study data were randomly split into a training set and a test set, with the test set composed of 20% of the observations. The 5-fold cross-validation was applied to fit the ML models in the training set. The popular approach, *RandomizedSearchCV* in *Scikit-learn*, was utilized to tune the hyperparameters. The hyperparameters were needed to define specific functions in each ML model to be learned. For example, the hyperparameters in penalized logistic regression included 'L–1', 'L–2' and 'elastic net' penalty. The hyperparameters in the random forest included the number of trees, the trees' maximum depth and the maximum number of variables at each split. The hyperparameters in the neural network included learning rate, layer sizes and activation function. A grid search with cross-validation was used to train different ML models in the training set to build better classifiers. The trained ML models were then fitted using the test set. The outcome variable was highly imbalanced: only 10% of individuals received at least 11 opioid prescriptions, while ~90% of individuals received less than 11 opioid prescriptions. The random oversample method was used in training models to mitigate the negative impact of the imbalanced problem on models' performance.

Statistical analysis

Differences in the categorical variables between the opioid high-frequency users and low-frequency users were tested via Pearson's chi-squared statistics. Differences in continuous variables were tested via *t*-tests. Each patient was assumed to have equal weight. The performance measure of the predictions included two widely used evaluation metrics: the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). The AUROC score represents the degree of separability of the model. The higher the AUROC score, the better the model at distinguishing between the two classes of the outcome variable. The receiver operating characteristic (ROC) curve plots the true positive rate in study outcomes against the false positive rate at various thresholds settings.[30] Like the AUROC, the higher the AUPRC score, the better the model at distinguishing between the two classes. The precision-recall curve (PRC) plots positive predictive values against true positive values.[31] To interpret AUROC and AUPRC, notice that the baseline score for AUROC is 0.5 – a random classifier will achieve an AUROC of 0.5; the baseline score for AUPRC is equal to the ratio of positive examples over total example.[32, 33] In this study, this ratio is the number of high-frequency opioid users over the total users and equal to 11.2%. One advantage of AUPRC over AUROC is that AUPRC is a more useful performance measure for imbalanced data when concerned with discovering positive cases.[34] For this study, the AUPRC was used as a baseline metric evaluating prediction performance. To determine the importance of the patient characteristics in explaining the opioid use frequency, the relative importance of the characteristics in the best performing model was computed based on the reduction in the Gini criterion used to select split points during the modelling process. The larger the reduction in the Gini index, the higher the relative importance of the variable. Data analysis codes were made available on the following website upon acceptance of the manuscript for publication: mtmstarvalue.uthsc.edu/codes.

Results

Descriptive statistics

Summary statistics of the characteristics of the study sample were analysed (Table 1). There were 7915 individuals in the sample, of which 11.2% (n = 888) were high-frequency opioid users. Comparing the low-frequency opioid users and high-frequency opioid users showed significant differences ($P < 0.01$) in all the characteristics except gender and census region. Compared with the low-frequency opioid users, high-frequency opioid users were more likely to be older, be non-Hispanic Whites, be unmarried, have earned lower educational degrees, have public insurance only, belong to lower-income categories, perceive their health status to be in the lower health categories, and have a higher number of chronic conditions.

Results from machine learning

Predictive performances of the different ML in predicting high-frequency opioid users were reported (Table 2). In terms of AUROC, the random forest achieved the highest score (0.7726), followed by gradient boosting (0.7679), support vector machine (0.7628), XGBoost (0.7563), penalized logistic regression (0.7537) and neural network (0.7530). Concerning AUPRC, the random forest achieved the highest score (0.2871), followed by gradient boosting (0.2846), neural network (0.2842), XGBoost (0.2740), penalized logistic regression (0.2665) and support vector machine (0.2659). The random forest turned out to be the best performing model in both AUROC and AUPRC measures of prediction performance. Moreover, the AUROC (0.7726) and AUPRC (0.2871) of random forest were much higher than the baseline scores of 0.5 and 0.11. Furthermore, the penalized logistic regression, usually chosen as the compare group, only performed slightly better than the neural network in AUROC and slightly better than the support vector machine in AUPRC. Curve plots of AUROC and AUPRC for the best performing model, the random forest, are depicted in Figures 1 and 2. In both Figures 1 and 2, the 'No Skill' line represented the baseline model (i.e. random guessing) with AUROC of 0.5 and AUPRC of 0.11. Both figures showed that random forest significantly improved the prediction performance relative to the random guessing.

To evaluate each variable's importance in explaining high-frequency opioid use in the best performing model (i.e. random forest), the relative importance of the top 10 variables was calculated (Table 3). The most influential feature in explaining the high-frequency opioid use was age, followed by the number of chronic conditions (5 or more), public insurance only, self-perceived health status (fair) and self-perceived health status (poor).

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

Table 1 Individual characteristics in the study population (number and frequency unless otherwise specified)

Characteristics	All (n = 7915)		High-frequency opioid users (n = 888; 11.2%)		Low-frequency opioid users (n = 7027; 88.8%)		P-value
	Number	%	Number	%	Number	%	
Predisposing factors							
Age, mean (SD)	53.5 (17.0)		58.0 (13.4)		52.9 (17.4)		<0.0001
Male	2937	37.1	335	37.7	2602	37.0	0.6856
Race/ethnicity							<0.0001
Non-Hispanic Whites	4678	59.1	582	65.5	4096	58.3	
Non-Hispanic Blacks	1467	18.5	160	18.0	1307	18.6	
Hispanics	1295	16.4	100	11.3	1195	17.0	
Non-Hispanic Asians	176	2.2	5	0.6	171	2.4	
Others	299	3.8	41	4.6	258	3.7	
Married	3770	47.6	382	43.0	3388	48.2	0.0035
Education > high school	2483	31.4	197	22.2	2286	32.5	<0.0001
Enabling factors							
Insurance type							<0.0001
Any private	4320	54.6	304	34.2	4016	57.2	
Public only	3256	41.1	553	62.3	2703	38.5	
No insurance	399	4.3	31	3.5	308	4.4	
Poverty category							<0.0001
Poor	1964	24.8	300	33.8	1664	23.7	
Low income	1286	16.3	192	21.6	1094	15.6	
Middle income	2159	27.3	215	24.2	1944	27.7	
High income	2326	29.4	152	17.1	2174	30.9	
Census region							0.0611
Northeast	1025	13.0	111	12.5	914	13.0	
Midwest	1866	23.6	214	24.1	1652	23.5	
South	3253	41.1	393	44.3	2860	40.7	
West	1771	22.4	170	19.1	1601	22.8	
Need factors							
Self-perceived health status							<0.0001
Excellent	788	10.0	20	2.3	768	10.9	
Very good	1875	23.7	92	10.4	1783	25.4	
Good	2648	33.5	252	28.4	2396	34.1	
Fair	1822	23.0	332	37.4	1490	21.2	
Poor	782	9.9	192	21.6	590	8.4	
Number of chronic conditions							<0.0001
≤1	2238	28.3	61	6.9	2177	40.0	
2–4	2775	35.1	255	28.7	2520	35.9	
≥5	2902	36.7	572	64.4	2330	33.2	

SD, standard deviation. Differences in categorical variables between opioid high-frequency users and low-frequency users were tested via Pearson's chi-squared tests. Difference in age was tested via t-test.

Table 2 Performance results of machine learning models in classifying patients on opioid use

Model	Evaluation metrics	
	AUROC	AUPRC
Penalized logistic regression	0.7537	0.2665
Support vector machine	0.7628	0.2659
Random forest	**0.7726**	**0.2871**
Neural network	0.7530	0.2842
Gradient boosting	0.7679	0.2846
XGBoost	0.7563	0.2740

AUROC is the area under the receiver operating characteristic curve while AUPRC is the area under the precision-recall curve. For both AUROC and AUPRC, the larger the score, the better the model's prediction performance. Random forest had the highest AUROC and AUPRC scores (bolded).

Discussion

The current opioid epidemic has prompted increased concerns over the opioid utilization patterns. This study makes a contribution to the literature related to opioid crisis since this is the first study employing several different ML models and logistic regression to predict high-frequency opioid utilization in a national representative sample. The focus on a higher number of opioid prescriptions for this study is vital since high-frequency opioid use tends to increase an individual's risk of opioid use disorder, overdose and death. In this study, predictors were selected according to the widely used Gelberg–Andersen Model. The study results revealed that ML models had better performance in predicting whether an individual lies in the upper 10% of the distribution of opioid prescriptions than logistic regression. The success of ML in this study indicates that researchers could use ML techniques to improve the prediction performance for high-frequency opioid use.

As documented in the literature, when data are imbalanced (i.e. low prevalence of outcome of interest), AUROC, a widely used performance measure for binary classifiers, may be misleading.[31] In this case, AUPRC is recommended as a more accurate evaluation measure of model performance.[32] AUROC and AUPRC were both reported in this study, but AUPRC was utilized as the baseline evaluation metric. After training each model using the training dataset, the results revealed

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

that ML techniques had acceptable performance in predicting high-frequency opioid users based on existing criteria for classifying ML model performance.[35] The random forest seems to be the best performing model since it achieved the highest values of both AUROC and AUPRC in the independent test dataset, outperforming the penalized logistic regression. However, the significance of differences between models was not tested, and it appears that the random forest model and gradient boosting exhibited especially similar performance.

Given the current opioid crisis, it is vital to identify the variables that play an essential role in predicting opioid prescriptions since this could assist the government and healthcare providers in targeting policies to prevent opioid overdose, misuse and death. In the best performing model (i.e. random forest), age, the number of chronic conditions (5 or more), public insurance only, and self-perceived fair and poor health status had relatively high predicting power in high-frequency opioid use. This result was expected since opioids have been widely considered the most effective drugs for treating chronic pain.[6] Older individuals with poor perceived health status are more likely to have chronic pain and receive more opioid prescriptions.[6] Previous literature also shows that individuals with public insurance are more likely to consume more opioids.[6] Therefore, healthcare providers

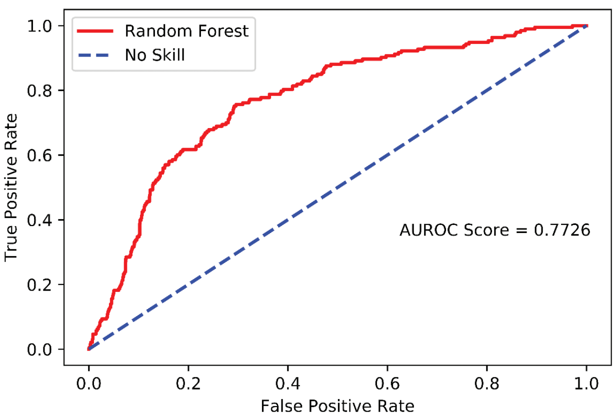

Figure 1 Performance for random forest in classifying patients on opioid use was measured using the area under the receiver operating characteristic curve (AUROC). 'No Skill' line represented the baseline model (i.e. random guessing) with AUROC of 0.5.

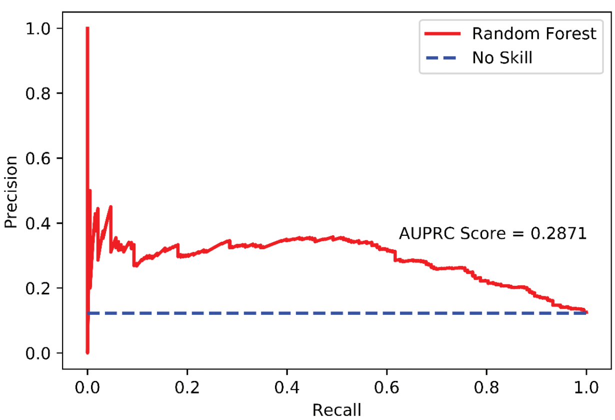

Figure 2 Performance for random forest in classifying patients on opioid use was measured using the area under the precision-recall curve (AUPRC). 'No Skill' line represented the baseline model (i.e. random guessing) with AUPRC of 0.11.

and government should be mindful that patient characteristics such as older age, with multiple chronic conditions, with public insurance only and with poorer self-perceived health status, have relatively high predictive power of high-frequency opioid use. Procedures may need to be set in place to use caution when prescribing and distributing prescription opioids if a patient possesses these factors.

This study has several limitations. First, a nationally representative sample of non-institutionalized civilians from MEPS was used; hence, the findings in this study cannot be generalized to other populations such as institutionalized civilians in the USA.[12, 36] The purpose of this study was not to develop a model that can be used in the future, but to determine whether ML could help to produce additional insights into the prediction of high-frequency opioid use. Second, even though five popular ML techniques in classifying binary outcomes were applied, the possibility that other available ML techniques could perform better than this study's best performing model cannot be ruled out. There is no consensus in the literature on the best choice of ML. However, various ML models were used for this study, and they provided very good prediction performance.[37] Third, the prediction performance of these ML techniques might be changed if a different set of predictors is chosen.[12] However, for this study, predictors were selected according to Gelberg–Andersen Model. These predictors were believed to be the most important factors affecting high-frequency opioid utilization. Finally, ML techniques require special software and skills, and the findings can be harder to interpret than traditional regression models. However, ML tools can provide valuable information for predicting opioid use. For example, while the difference between AUPRC of logistic regression and the random forest was only 0.0206 in this study, the 7.7% difference between these models [(0.2871–0.2665)/0.2665] suggests a

Table 3 Relative importance of the patient characteristics in the random forest

Variables	Relative importance	Rank
Age	0.2273	1
Number of chronic conditions (≥5)	0.2102	2
Insurance type (public insurance only)	0.1007	3
Self-perceived health status (fair)	0.0704	4
Self-perceived health status (poor)	0.0538	5
Number of chronic conditions (2–4)	0.0473	6
Self-perceived health status (very good)	0.0458	7
Poverty category (high income)	0.0350	8
Race/ethnicity (Hispanics)	0.0307	9
Education > high school	0.0226	10
Male	0.0182	11
Married	0.0177	12
Self-perceived health status (good)	0.0162	13
Race/ethnicity (non-Hispanic Asians)	0.0151	14
Poverty category (low income)	0.0137	15
Poverty category (middle income)	0.0126	16
Census region (South)	0.0122	17
Census region (West)	0.0119	18
Race/ethnicity (non-Hispanic Blacks)	0.0110	19
Insurance type (no insurance)	0.0107	20
Census region (Midwest)	0.0096	21
Race/ethnicity (Others)	0.0072	22

Random forest was found to be the best performing model in terms of AUROC and AUPRC. Relative importance of each variable was computed based on the reduction in the Gini criterion used to select split points during the modelling process. The larger the reduction in the Gini index, the higher the relative importance of the variable.

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

meaningful improvement. Future studies may further test the significance of the differences between various ML models.

Despite the limitations listed above, this study makes a contribution to the literature as this is the first study employing several different ML models and logistic regression to predict high-frequency opioid utilization in a national representative sample. The results of this study illustrate that random forest performed best in two popular evaluation metrics: AUROC and AUPRC. Further, it identified several influential variables in predicting the frequency of opioid prescription use. These ML models were found to have a more robust and superior prediction performance than logistic regression. These models could be used as valuable tools to accurately and efficiently identify individuals potentially at high risk of receiving too many opioid prescriptions. The study results demonstrate that ML can be a promising and powerful technique in health outcome predictions.

Conclusions

This study applied five different ML techniques and penalized logistic regression to predict the frequency of opioid prescriptions. Two ML methods, random forest and gradient boosting showed the best and similar performance, with random forest achieving the highest scores. Age, the number of chronic conditions (5 or more), public insurance only and self-perceived health status were found to have notable predicting power in the random forest. These findings provide a potential opportunity for healthcare providers to identify the frequent use of opioid prescriptions. Future research related to predictions in medical and health outcomes should consider comparing various ML techniques over conventional statistical methods.

Acknowledgements

The authors would like to acknowledge the assistance with the formatting and submission of the final manuscript from Lorraine Todor, a PharmD student at the University of Tennessee Health Science Center College of Pharmacy.

Author Contributions

All authors have contributed to this study and all authors reviewed and approved the final version of the manuscript. SZ and JW participated in the study concept and design, analysis and interpretation of data and writing (original draft and editing). JB and YC participated in the study concept and design, interpretation of data and writing (original draft and editing).

Funding

None.

Conflict of Interest

J.W. received funding from AbbVie, Curo, Bristol Myers Squibb, Pfizer, National Institutes of Health and Pharmaceutical Research and Manufacturers of America (PhRMA) and serves on Health Outcomes Research Advisory Committee of the PhRMA Foundation. Other authors declare that they have no conflict of interest.

Data Availability

Data are available via the Agency for Healthcare Research and Quality at the following website: https://www.meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp.

References

1. CDC Injury Center. U.S. Opioid Dispensing Rate Maps. https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html (23 September 2020, date last accessed).
2. Division (DCD) DC. 5-Point Strategy to Combat the Opioid Crisis. https://www.hhs.gov/opioids/about-the-epidemic/hhs-response/index.html (23 September 2020, date last accessed).
3. CDC Injury Center. *Overdose Death Maps*. https://www.cdc.gov/drugoverdose/data/prescribing/overdose-death-maps.html (23 September 2020, date last accessed).
4. Buchmueller T, Carey C. The effect of prescription drug monitoring programs on opioid utilization in Medicare. *Am Econ J Econ Policy* 2018; 10: 77–112. http://doi.org/10.1257/pol.20160094
5. Daubresse M, Saloner B, Pollack HA *et al*. Non-buprenorphine opioid utilization among patients using buprenorphine. *Addiction* 2017; 112: 1045–53. http://doi.org/10.1111/add.13762
6. Herzig SJ, Rothberg MB, Cheung M *et al*. Opioid utilization and opioid-related adverse events in nonsurgical patients in US hospitals. *J Hosp Med* 2014; 9: 73–81. http://doi.org/10.1002/jhm.2102
7. Kalakoti P, Hendrickson NR, Bedard NA *et al*. Opioid utilization following lumbar arthrodesis: trends and factors associated with long-term use. *Spine (Phila Pa 1976)* 2018; 43: 1208–16. http://doi.org/10.1097/BRS.0000000000002734
8. Moyo P, Simoni-Wastila L, Griffin BA *et al*. Impact of prescription drug monitoring programs (PDMPs) on opioid utilization among Medicare beneficiaries in 10 US States. *Addiction* 2017; 112: 1784–96. http://doi.org/10.1111/add.13860
9. Pezalla EJ, Rosen D, Erensen JG *et al*. Secular trends in opioid prescribing in the USA. *J Pain Res* 2017; 10: 383–7. http://doi.org/10.2147/JPR.S129553
10. Politzer CS, Kildow BJ, Goltz DE *et al*. Trends in opioid utilization before and after total knee arthroplasty. *J Arthroplasty* 2018; 33: S147–53.e1. http://doi.org/10.1016/j.arth.2017.10.060
11. Prentice HA, Inacio MCS, Singh A *et al*. Preoperative risk factors for opioid utilization after total hip arthroplasty. *J Bone Jt Surg* 2019; 101: 1670–8. http://doi.org/10.2106/JBJS.18.01005.
12. Han DH, Lee S, Seo DC. Using machine learning to predict opioid misuse among U.S. adolescents. *Prev Med* 2020; 130: 105886. http://doi.org/10.1016/j.ypmed.2019.105886
13. Badger J, LaRose E, Mayer J *et al*. Machine learning for phenotyping opioid overdose events. *J Biomed Inform* 2019; 94: 103185. http://doi.org/10.1016/j.jbi.2019.103185
14. Che Z, St Sauver J, Liu H *et al*. Deep learning solutions for classifying patients on opioid use. *AMIA Annu Symp Proc* 2017; 2017: 525–34.
15. Dong X, Rashidian S, Wang Y *et al*. Machine learning based opioid overdose prediction using electronic health records. *AMIA Annu Symp Proc* 2019; 2019: 389–98.
16. Ellis RJ, Wang Z, Genes N *et al*. Predicting opioid dependence from electronic health records with machine learning. *BioData Min* 2019; 12: 3. http://doi.org/10.1186/s13040-019-0193-0
17. Karhade AV, Ogink PT, Thio QCBS *et al*. Machine learning for prediction of sustained opioid prescription after anterior cervical discectomy and fusion. *Spine J* 2019; 19: 976–83. http://doi.org/10.1016/j.spinee.2019.01.009
18. Karhade AV, Ogink PT, Thio QCBS *et al*. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J* 2019; 19: 1764–71. http://doi.org/10.1016/j.spinee.2019.06.002
19. Katakam A, Karhade AV, Schwab JH *et al*. Development and validation of machine learning algorithms for postoperative opioid prescriptions after TKA. *J Orthop* 2020; 22: 95–9. http://doi.org/10.1016/j.jor.2020.03.052
20. Lee S, Wei S, White V *et al*. Classification of opioid usage through semi-supervised learning for total joint replacement patients. *IEEE J Biomed Health Inform* 2021; 25: 189–200. http://doi.org/10.1109/JBHI.2020.2992973

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023

21. Lo-Ciganic WH, Huang JL, Zhang HH *et al*. Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Netw Open* 2019; 2: e190968. http://doi.org/10.1001/jamanetworkopen.2019.0968
22. Lo-Ciganic WH, Huang JL, Zhang HH *et al*. Using machine learning to predict risk of incident opioid use disorder among fee-for-service Medicare beneficiaries: a prognostic study. *PLoS One* 2020; 15: e0235981. http://doi.org/10.1371/journal.pone.0235981
23. Workman TE, Shao Y, Kupersmith J *et al*. Explainable deep learning applied to understanding opioid use disorder and its risk factors. In: 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019, 4883–8.
24. Wiles NJ, Zammit S, Bebbington P *et al*. Self-reported psychotic symptoms in the general population: results from the longitudinal study of the British National Psychiatric Morbidity Survey. *Br J Psychiatry* 2006; 188: 519–26. http://doi.org/10.1192/bjp.bp.105.012179
25. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 2016; 5: 221–32. http://doi.org/10.1007/s13748-016-0094-0.
26. Agency for Healthcare Research and Quality. *Medical Expenditure Panel Survey Background*. https://www.meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp (23 September 2020, date last accessed).
27. CDC Injury Center. *CDC Compilation of Benzodiazepines, Muscle Relaxants, Stimulants, Zolpidem, and Opioid Analgesics with Oral Morphine Milligram Equivalent Conversion Factors (2017 Version)*. https://www.cdc.gov/drugoverdose/data-files/CDC_Oral_Morphine_Milligram_Equivalents_Sept_2017. xlsx (23 September 2020, date last accessed).
28. Gelberg L, Andersen RM, Leake BD. The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. *Health Serv Res* 2000; 34: 1273–302.
29. Chronic Condition Indicator (CCI) for ICD-10-CM (beta version). https://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp (23 September 2020, date last accessed).
30. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 1997; 30: 1145–59. http://doi.org/10.1016/S0031-3203(96)00142-2.
31. Davis J, Goadrich M. The relationship between precision-recall and Roc curves. In: Proceedings of the 23rd International Conference on Machine Learning - ICML'06, Pittsburgh, PA, USA, 25–29 June 2006. New York, NY: ACM Press, 2006, 233–40.
32. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015; 10: e0118432. http://doi.org/10.1371/journal.pone.0118432
33. Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 2015; 68: 855–9. http://doi.org/10.1016/j.jclinepi.2015.02.010
34. Chiew CJ, Liu N, Wong TH *et al*. Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. *Ann Surg* 2020; 272: 1133–9. http://doi.org/10.1097/SLA.0000000000003297
35. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd edn., Chapter 5. New York, NY: John Wiley & Sons, 2000, 177.
36. Wang J, Munshi KD, Hong SH. Racial and ethnic disparities in influenza vaccinations among community pharmacy patients and non-community pharmacy respondents. *Res Social Adm Pharm* 2014; 10: 126–40. http://doi.org/10.1016/j.sapharm.2013.04.011
37. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB *et al*. Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. *PLoS One* 2016; 11: e0155705. http://doi.org/10.1371/journal.pone.0155705

Downloaded from https://academic.oup.com/jphsr/article/12/4/502/6402781 by guest on 20 January 2023