OPEN ACCESS

Machine learning approach to predicting persistent opioid use following lower extremity joint arthroplasty

Rodney A Gabriel,[1] Bhavya Harjai,[1] Rupa S Prasad,[1] Sierra Simpson,[2] Iris Chu,[1] Kathleen M Fisch,[3] Engy T Said [1,4]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/rapm-2021-103299).

[1]Anesthesiology, University of California San Diego, La Jolla, California, USA
[2]Psychiatry, University of California San Diego, La Jolla, California, USA
[3]Center for Computational Biology and Bioinformatics, University of California San Diego, La Jolla, California, USA
[4]University of California San Diego, San Diego, California, USA

Correspondence to
Dr Rodney A Gabriel, Anesthesiology, University of California San Diego, La Jolla, California, USA; ragabriel@health.ucsd.edu

Received 31 October 2021
Accepted 18 January 2022
Published Online First 3 February 2022

ABSTRACT

Background The objective of this study is to develop predictive models for persistent opioid use following lower extremity joint arthroplasty and determine if ensemble learning and an oversampling technique may improve model performance.

Methods We compared various predictive models to identify at-risk patients for persistent postoperative opioid use using various preoperative, intraoperative, and postoperative data, including surgical procedure, patient demographics/characteristics, past surgical history, opioid use history, comorbidities, lifestyle habits, anesthesia details, and postoperative hospital course. Six classification models were evaluated: logistic regression, random forest classifier, simple-feedforward neural network, balanced random forest classifier, balanced bagging classifier, and support vector classifier. Performance with Synthetic Minority Oversampling Technique (SMOTE) was also evaluated. Repeated stratified k-fold cross-validation was implemented to calculate F1-scores and area under the receiver operating characteristics curve (AUC).

Results There were 1042 patients undergoing elective knee or hip arthroplasty in which 242 (23.2%) reported persistent opioid use. Without SMOTE, the logistic regression model has an F1 score of 0.47 and an AUC of 0.79. All ensemble methods performed better, with the balanced bagging classifier having an F1 score of 0.80 and an AUC of 0.94. SMOTE improved performance of all models based on F1 score. Specifically, performance of the balanced bagging classifier improved to an F1 score of 0.84 and an AUC of 0.96. The features with the highest importance in the balanced bagging model were postoperative day 1 opioid use, body mass index, age, preoperative opioid use, prescribed opioids at discharge, and hospital length of stay.

Conclusions Ensemble learning can dramatically improve predictive models for persistent opioid use. Accurate and early identification of high-risk patients can play a role in clinical decision making and early optimization with personalized interventions.

To cite: Gabriel RA, Harjai B, Prasad RS, *et al*. *Reg Anesth Pain Med* 2022;**47**:313–319.

INTRODUCTION

The USA is in the midst of an opioid epidemic, with the number of opioid-related deaths having risen sixfold since 1999.[1] Surgery is a risk factor for chronic opioid use, with up to a 3% incidence of previously opioid naïve patients continuing to use opioids for more than 90 days after a major elective surgery.[2] Joint arthroplasty surgery is one of the most common surgical procedures performed in the USA, with over 1 million knee and hip replacements occurring annually.[3] Despite improvements in multimodal analgesia, 10%–40% of patients undergoing lower extremity joint arthroplasty still develop chronic postsurgical pain.[4]

Several studies have investigated risk factors for persistent opioid use following total knee and hip arthroplasty.[5 6] Preoperative opioid use has consistently been found to increase the risk of postoperative chronic opioid use.[7–11] Other patient characteristics associated with increased risk of chronic opioid use include history of depression, higher baseline pain scores, younger age, and female sex.[7–10] Additional research is needed to develop tools to more accurately predict patients at highest risk postoperatively. Identification of at-risk patients prior to hospital discharge allows time for the formulation of a pre-emptive individualized pain management plan. Novel modalities exist that may potentially help reduce persistent opioid use, including peripheral neuromodulation, cryoneurolysis, and transitional pain clinics, however, due to limited resources, it may not be realistic to offer such modalities to every patient.[12] Developing accurate predictive models will help better allocate these resources.

Limited research exists investigating the utility of machine learning in predicting persistent opioid use—defined as continued opioid use more than 3 months after surgery. The primary objective of this study is to develop machine learning-based predictive models predicting persistent opioid use. We will incorporate data from the entire acute perioperative period (preoperative, intraoperative, and postoperative variables) so that identification of high-risk patients can occur prior to hospital discharge. Furthermore, machine learning algorithms are typically evaluated by their predictive accuracy; however, when data is imbalanced (ie, large difference in rates of positive vs negative outcomes), model performance can be biased and problematic. Given that most patients undergoing joint arthroplasty do not develop persistent opioid use, we know that such datasets will be imbalanced. Therefore, to optimize our machine learning algorithms, we also applied an oversampling technique to balance the dataset using Synthetic Minority Oversampling Technique (SMOTE), which has been shown to help improve accuracy of models without biasing the study outcome.[13]

METHODS

Study sample

The informed consent requirement was waived. Data from all patients that underwent elective hip

or knee arthroplasty from 2016 to 2019 were extracted from the electronic medical record (EMR) database. Emergent cases, bilateral joint arthroplasty, hemiarthroplasty, and unicompartmental knee arthroplasty were excluded from the analysis. The manuscript adheres to the applicable EQUATOR guidelines for observational studies.

Primary objective and data collection

The primary outcome measurement was persistent opioid use, defined as patient reporting to use opioids after a 3-month postoperative cut-off, up to 6 months. The outcome data were extracted from the EMR system from: (1) orthopedic surgery postoperative follow-up note (at 3–6 months following surgery). In these notes, the surgeons routinely describe continued use of opioids or pain; (2) primary care physician or pain specialist notes (at 3–6 months postoperatively); (3) scanned documents from providers outside of our healthcare system (at 3–6 months postoperatively); and/or (4) active opioid prescription that is filled during this time period (captured in the EMR). We evaluated six different classification models: logistic regression, random forest classifier, simple-feedforward neural network, balanced random forest classifier, balanced bagging classifier, and support vector classifier. In addition, we evaluated model performance with SMOTE.

The covariates included in the models were: surgical procedure (total hip arthroplasty (posterolateral approach vs. anterior approach), total knee arthroplasty, revision total hip arthroplasty, and revision total knee arthroplasty), age, sex, body mass index, English as a primary language, preoperative opioid use, previous joint replacement surgery, osteoarthritis severity in the operative limb, hypertension, coronary artery disease, chronic obstructive pulmonary disease, asthma, obstructive sleep apnea, diabetes mellitus (non-insulin vs insulin-dependence), psychiatric history (anxiety and/or depression), active alcohol history (defined as ≥2 drinks per day), active smoking history, active marijuana use, use of perioperative regional nerve block, primary anesthesia type (neuraxial vs general anesthesia), intraoperative ketamine use (yes or no), opioid use on postoperative day 1 (measured in intravenous morphine equivalents (MEQ)), amount of prescription opioids given at discharge (MEQs), and hospital length of stay (days) (online supplemental table 1). To measure the amount of prescription opioids given at discharge, we defined this as number of pills multiplied by the opioid in MEQ. No data were missing.

Statistical analysis

Python (V.3.7.5) was used for all statistical analysis. First, the cohort was divided into training and test data sets, reflecting an 80:20 split using a randomized splitter—the 'train_test_split' method from the sci-kit learn library[12]—thus, any patients present in the test set were automatically removed from the training set. We developed each machine learning model using the same training set (with or without SMOTE) and tested its performance on the same test set (measuring F1 score, accuracy, recall, precision, and the area under the curve (AUC) for the receiver operating characteristics curve (ROC)). To perform a more robust evaluation of the models, we then calculated the average F1 score, accuracy, recall, precision, and AUC using stratified K-fold cross-validation (described below) (figure 1).

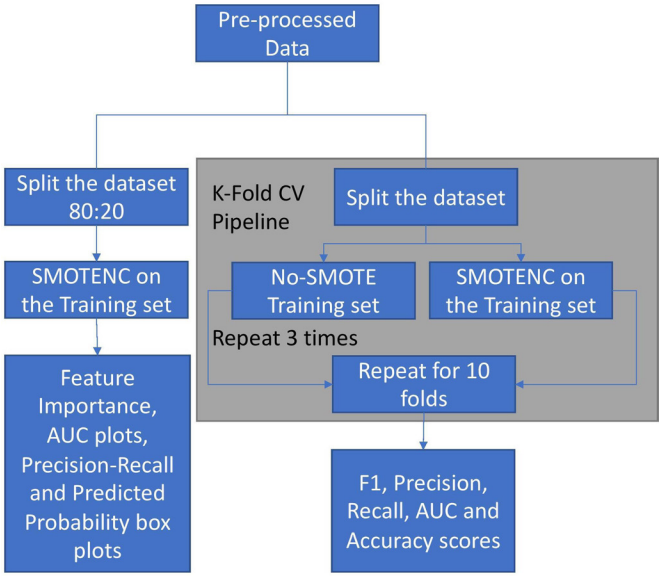

Figure 1 Analysis pipeline. AUC, area under the curve; CV, cross-validation; SMOTE, Synthetic Minority Oversampling Technique.

Data balancing

SMOTE for Nominal and Continuous algorithm—implemented using the 'imblearn' library (https://imbalanced-learn.org/stable/)—was used to create a balanced class distribution and was first described by Chawla *et al*.[13] Imbalanced data may be particularly difficult for predictive modeling due to the uneven classification of data. A balanced dataset would have minimal difference in positive and negative outcomes. However, if the difference is large, it is considered unbalanced. SMOTE is a statistical technique that increases the number of cases in a dataset to balance it—it does this by increasing new instances from minority cases, while not affecting the number of majority cases. This algorithm takes samples of the feature space for each target class and five of its nearest neighbors, and then generates new cases that combines features of the target case with features of its nearest neighbors. This method increases the percentage of the minority cases in the dataset. SMOTE was only applied to our training sets and we did not oversample the testing set, thus maintaining the natural outcome frequency.

Machine learning models

We evaluated six different classification models: logistic regression, random forest classifier, simple-feed-forward neural network, balanced random forest classifier, balanced bagging classifier, and support vector classifier. For each, we also compared the use of oversampling the training set via SMOTE vs no SMOTE. For each model, all features were included as inputs. Multivariable logistic regression—This is a statistical model that asserts a binary outcome based on the weighted combination of the underlying independent variables. We tested an L2-penalty-based regression model without specifying individual class weights. This model provides a baseline score and helps make the case for improvement over the evaluation metrics. Random forest classifier—We developed a random forest classifier with 1000 estimators, and the criterion for the split was set to the Gini impurity. The Gini impurity is calculated in which C is total number of classes and p(i) is the

C

$$G = i = \sum p(i) * (1 - p(i))$$

1

Random forest is a technique that combines the predictions from multiple machine learning algorithms together to make

more accurate predictions than any individual model.[14] The random forest is a robust and reliable non-parametric supervised learning algorithm that acts as a means to test the further improvement in the metrics and provide the feature importance of the dataset. Balanced random forest classifier—This is an implementation of the random forest, which randomly under-samples each bootstrap to balance it. The model was built using 1000 estimators and the default values provided in the imblearn package. The sampling strategy was set to 'auto', which is equivalent to 'not minority'. Other parameters were kept the same as our random forest classifier. Balanced bagging classifier—Another way to ensemble models is bagging or bootstrap-aggregating. Bagging methods build several estimators on different randomly selected subsets of data. Unlike random forests, bagging models are not sensitive to the specific data on which they are trained. They would give the same score even when trained on a subset of the data. Bagging classifiers are also generally more immune to overfitting. We built a balanced bagging classifier using the imblearn package, where the number of estimators was set to 1000 where replacement was allowed, and the sampling strategy was set to auto, which is equivalent to 'not minority,' and, thus, does not resample the minority class. Multilayer perceptron neural network—Using sci-kit learn's 'MLPClassifier,' we built a basic shallow feed-forward network with two hidden layers and ten neurons in each hidden layer. The activation function was set to the rectified linear unit function and the net was trained for a maximum of 700 iterations. The other parameters remained default as implemented in the sci-kit learn package. Support vector classifier—A support vector classifier tries to find a hyperplane decision boundary that best splits the data into the required number of classes. It plots each data item as a point in an n-dimensional space (n being the number of features), then finds a hyperplane that separates the classes. We developed a modification of support vector that weighs the margin proportional to the class importance, or is a cost-sensitive support vector classifier by choosing the gamma value as 'scale,' while defining the classifier and assigning 'balanced' to the class-weight parameter.

K-folds cross-validation

To perform a more robust evaluation of our models, we implemented repeated stratified-K-fold cross-validation to observe the accuracy, precision, recall, F1-score, and AUC scores, for 10 splits and 3 repeats. For each iteration, the dataset was split into 10-fold, where 1-fold served as the test set, and the remaining nine sets served as the training set. The model was built on the training set. In the case when SMOTE was used, only the training set was oversampled. This was repeated until all folds had the opportunity to serve as the test set. This was then repeated three times. For each iteration, our performance metrics were calculated on the test set. The average of each performance metric was calculated thereafter.

Performance metrics

Accuracy

Accuracy is the ratio of the correct predictions to the total number of data points present in the test set. This is an important metric used to evaluate the number of points the model predicted incorrectly, but does not tell us the full extent of the models' performance.

Accuracy=true positive +true negative/(true positive +true negative +false positive +false negative)

Precision

Precision is defined as a metric that quantifies the number of correct predictions made by the model. In a way, precision calculates the accuracy of the minority class. Formally, it is defined as the ratio of the True Positive samples to the Sum of the True Positive and False Positive samples.

Precision=true positive/(true positive +false positive).

Recall

Recall quantifies the number of correct positive predictions made from all the positive predictions that could have been made. It serves as an indication of missed positive predictions. It is formally defined as the ratio of true positives over the sum of true positives and false negatives.

Recall=true positive/(true positive +false negative)

F1-Score

This is a version of the Fβ-metric where we provide equal weight to the Precision and Recall scores. F1 score is formally equal to the harmonic mean of Precision and Recall and this provides a way to combine both into a single metric. This is the most valuable metric to analyze a classification task and thus, is the most significant metric of our analysis.[15 16]

F1 Score=2 × precision ×recall/(precision +recall).

Area under the curve

We calculated the AUC for the ROC curve. The ROC curve was developed by plotting the fraction of true positive rate vs the false positive rate and represents sensitivity versus one minus the specificity in the curve. The AUC summarizes the whole curve in just a number, ranging from 0 to 1, 1 being the best possible score to achieve.

RESULTS

Study population

Initially, there were 1094 patients, but after exclusion, there were 1042 patients in the final analysis, of which 242 (23.2%) required persistent opioid use after 3–6 months following surgery. The cohort of patients that did not have persistent opioid use had higher proportion of hip arthroplasties and severe osteoarthritis of the surgical joint. Those that did have persistent opioid use tended to be younger, received intraoperative ketamine, consumed more opioids postoperative day 1, had longer hospital length of stay, and had higher proportions of substance abuse, congestive heart failure, chronic obstructive pulmonary disease, and anxiety/depression (table 1).

We initially split the data into a training and test set and used SMOTE to oversample the training set for each model. When SMOTE was applied on the training set, there was a 1:1 ratio of positive and negative classes (table 2), in which there was a ∼3.2 times increase in positive classes (persistent opioid use).

The AUC of the ROC for the logistic regression model was 0.72, while each ensemble learning approach (ie, random forest, balanced bagging, and balanced random forest) had an AUC of 0.95 (figure 2). We generated box plots (figure 3) illustrating the quantiles of the probability scores generated from each machine learning model when validated on the test set. In the test set (n=219), there were 49 (22.4%) patients who developed persistent opioid use. When modeling development of persistent opioid use, the ensemble learning approaches correctly predicted patients who had probability scores≥0.5 more often compared with the other models (figure 3A).

Table 1 Patient characteristics in the two cohorts

	No persistent opioid use		Persistent opioid use		
	N	%	N	%	P value
Total	800		242		
Surgical procedure					<0.0001
Hip arthroplasty, posterolateral approach	224	28.0	26	10.7	
Hip arthroplasty, anterior approach	161	20.1	55	22.7	
Revision hip arthroplasty	42	5.3	18	7.4	
Total knee arthroplasty	315	39.4	111	45.9	
Revision knee arthroplasty	58	7.3	32	13.2	
Age (years), mean (SD)	68.7(10.5)		64.9(11.2)		<0.0001
Male sex	271	33.9	77	31.8	0.61
Body mass index (kg/m^2), mean (SD)	29.1 (5.0)		30.1 (5.6)		0.01
English-speaking	746	93.3	220	90.9	0.28
Preoperative opioid use	83	10.4	111	45.9	<0.0001
Previous joint replacement surgery	330	41.3	114	47.1	0.12
Comorbidities					
Hypertension	473	59.1	145	59.9	0.88
Coronary artery disease	75	9.4	34	14.0	0.05
Congestive heart failure	20	2.5	19	7.9	0.0003
Chronic obstructive pulmonary disease	40	5.0	28	11.6	0.0005
	No persistent opioid use		Persistent opioid use		
Comorbidities continued	n	%	n	%	P value
Asthma	77	9.6	23	9.5	0.99
Obstructive sleep apnea	132	16.5	54	22.3	0.05
Diabetes mellitus	119	14.9	50	20.7	0.04
Insulin-dependent diabetes mellitus	26	3.3	13	5.4	0.18
Anxiety/depression	208	26.0	92	38.0	0.0004
Dementia	2	0.3	0	0.0	0.99
Renal insufficiency	60	7.5	13	5.4	0.32
Frequent alcohol use (≥1 drink per day)	387	48.4	122	50.4	0.63
Active smoker	222	27.8	74	30.6	0.13
Frequent marijuana use (≥once/week)	25	3.1	23	9.5	<0.0001
Active illicit drug use	9	1.1	12	5.0	0.0005
Severe osteoarthritis of the surgical joint	486	60.8	122	50.4	0.005
Fracture	8	1.0	18	7.4	<0.0001
Intraoperative					
Primary anesthetic: neuraxial	460	57.5	133	55.0	0.53
Peripheral nerve block performed	349	43.6	133	55.0	0.002
Intraoperative dexmedetomidine	181	22.6	52	21.5	0.78
Intraoperative ketamine	48	6.0	32	13.2	0.0004
Postoperative					
POD 1 opioid consumption (MEQ)	13.2(20.3)		24.7 (22.6)		<0.0001
Hospital length of stay (days), mean (SD)	2.3 (1.6)		3.3 (2.4)		<0.0001
Discharge medications					
Total opioids prescribed at discharge (MEQ)	235.4 (176.5)		234.2 (172.1)		0.92
Oxycodone prescribed	685	85.6	195	80.6	0.07
Morphine prescribed	45	5.6	42	17.4	<0.0001
Hydromorphone prescribed	14	1.8	12	5.0	0.01
Tramadol prescribed	124	15.5	30	12.4	0.28
Hydrocodone prescribed	22	2.8	0	0.0	0.02

χ^2 test was used to compare categorical variables.
Student's t-test was used to compare continuous variables.
MEQ, intravenous morphine equivalent; POD, postoperative day.

Based on the balanced random forest model, we reported the features contributing most to the model (figure 4). The most important features contributing to prediction with the balanced bagging approach were (in descending order) postoperative day 1 opioid consumption, body mass index, age, preoperative opioid use, opioids prescribed at discharge, hospital length of stay, and the surgical procedure.

Table 2 Sample size distribution of positive (persistent opioid use) and negative classes (no persistent opioid use) in the original training set (80% of total sample) vs SMOTE dataset

Dataset	Positive classes	Negative classes	Total
Original training set	200	634	834
SMOTE	634	634	1268

SMOTE, Synthetic Minority Oversampling Technique.

Performance metrics calculated from K-folds cross-validation

Without SMOTE: We then calculated the average F1 score, accuracy, precision, recall, and AUC from k-folds cross-validation among all models, first without using SMOTE (table 3). When using logistic regression, the F1 score and AUC was 0.47 and 0.79, respectively. In comparison, the best performing model was balanced bagging classifier with the F1 score of 0.80 and AUC of 0.94.

With SMOTE: When SMOTE was used to oversample the test sets, the F1 score for most models improved (table 3, values in green and red font signify improvement or decreased performance, respectively, for the given metric when SMOTE was applied). For example, for the balanced bagging classifier, the F1 score improved from 0.80 (no SMOTE) to 0.84 (with SMOTE). For this model type, the AUC improved from 0.94 (no SMOTE) to 0.96. However, there were cases where accuracy, precision, recall, and AUC had decreased performance when SMOTE was applied. For example, the AUC of the multilayer perceptron went from 0.78 (no SMOTE) to 0.76 (with SMOTE). In addition, we report performance metrics of each machine learning model when different ratios of positive to negative classes were applied to SMOTE (online supplemental table 2).

DISCUSSION

We demonstrated that using an ensemble machine learning approach, combined with an oversampling technique of the training set, can improve the prediction of persistent opioid use following joint replacement. In our study population, we found the prevalence of persistent opioid use to be 23.7%. While there are several interventions that may potentially reduce this incidence—such as transitional pain clinic, peripheral neuromodulation, cryoneurolysis—these additional therapies may not realistically be applied to every patient. Thus, improving the ability to risk stratify and identify the at-risk population at time of hospital discharge is of utmost importance.

With expanding surveillance and access to electronic health data, methodologies in artificial intelligence are becoming pertinent in prediction analysis. Using the same dataset, we can improve our ability to predict outcomes by applying different types of machine learning approaches. Such practice should be applied more often

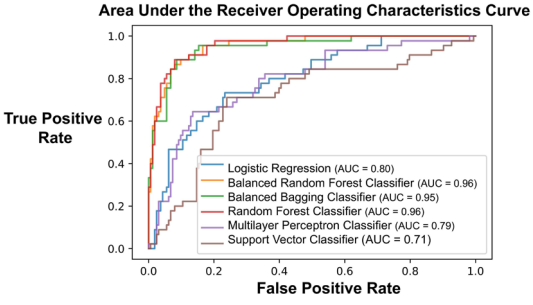

Figure 2 Area under the receiver operating characteristics curve for six separate models—logistic regression, multilayer perceptron neural network classifier, balanced random forest, balanced bagging classifier, random forest classifier, and support vector classifier. AUC, area under the curve.

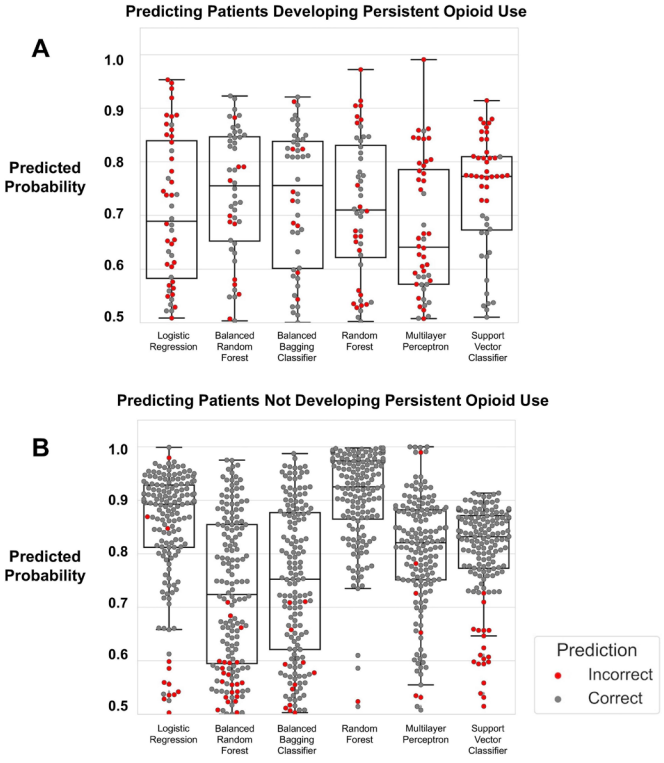

Figure 3 Box plot illustrating the distribution of probability scores generated by each type of machine learning model. Each dot represents a patient. (A) All models predicting development of persistent opioid use. Those with probability score ≥0.5 is assumed to mean subject is high risk for this outcome. A gray dot signifies that the model correctly classified the patient as developing the outcome, whereas a red dot signifies inaccurate classification. (B) All models predicting that patient will not develop persistent opioid use. A gray dot signifies that the model correctly classified the patient has not developing persistent opioid use. A red dot signifies inaccurate classification.

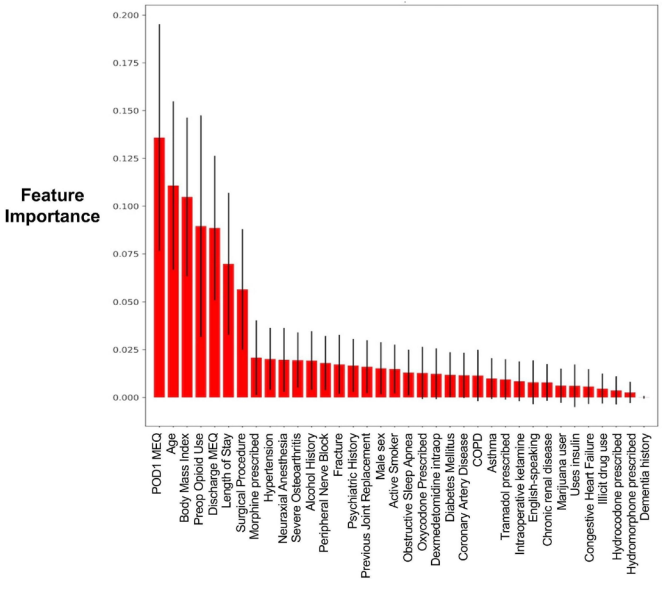

Figure 4 Opioid feature importance graph of 36 features based on the balanced random forest approach. COPD, chronic obstructive pulmonary disease; MEQ, intravenous morphine equivalents; POD1, postoperative day 1.

Table 3 Performance metrics on each machine learning approach with versus without using Synthetic Minority Oversampling Technique

	F1		Accuracy		Precision		Recall		AUC	
Machine learning approach	No SMOTE	SMOTE	No SMOTE	SMOTE	No SMOTE	SMOTE	No SMOTE	SMOTE	No SMOTE	SMOTE
Logistic regression	0.473	0.542	0.806	0.749	0.643	0.473	0.379	0.644	0.794	0.766
Balanced random forest classifier	0.747	0.847	0.863	0.933	0.656	0.891	0.874	0.813	0.936	0.959
Balanced bagging classifier	0.803	0.841	0.901	0.931	0.752	0.887	0.869	0.806	0.942	0.959
Random forest classifier	0.797	0.847	0.919	0.932	0.934	0.884	0.701	0.818	0.957	0.959
Multilayer perceptron classifier	0.399	0.505	0.802	0.712	0.690	0.440	0.301	0.638	0.777	0.759
Support vector classifier	0.475	0.449	0.724	0.653	0.436	0.603	0.531	0.603	0.727	0.707

Values in green font signify improvement in given metric when SMOTE is used. Values in red font signify decrease in performance of given metric when SMOTE is used.
AUC, area under the curve; SMOTE, Synthetic Minority Oversampling Technique.

in the healthcare setting, given the exponential increase in EMR data acquisition and availability. A basic approach to identifying associations between patient characteristics and outcomes includes regression techniques. In our study, we showed that when using an ensemble learning approach, the prediction of persistent opioid use is improved compared with regression, which have a few limitations: most notably, they only capture linear relationships between features and the outcomes.

Ensemble learning is beneficial because it leverages multiple learning algorithms and techniques to better predict an outcome. However, it requires diversity within the sample and between the models. To accomplish this, methods such as bagging, the use of different classifiers, and oversampling can be used to generate diversity and class balance within a given dataset. Often, data are lost due to undersampling, therefore, we can apply techniques that over sample to match the samples from the minority class to the majority class. Electronic health data are well known for class imbalance when a given outcome may not be prevalent within a population.[17] Two methods to overcome this problem include random oversampling and synthetic generation of minority class data by SMOTE. Instead of replicating random data from the minority class, SMOTE uses a nearest-neighbor approach to generate synthetic data to reduce the class imbalance. We demonstrated that when we apply SMOTE to our training set, we can improve model performance with our ensemble learning techniques. This is likely because the generation of synthetic points in the training dataset may allow for better validation of the test dataset. However, there are some limitations with SMOTE; namely, it may have limitations with high-dimensional data (and thus introduce additional noise when oversampling the minority class), such as the case in study populations with increased heterogeneity and features.[18] As we fine-tune our predictive ability further, we can optimize our ability to use healthcare resources efficiently to manage patient care via a personalized medicine approach.

Using the balanced random forest model's feature importance plot, we identified six variables to be the most important predictors in our models for persistent postoperative opioid use following joint replacement. These factors include: postoperative day 1 opioid use, body mass index, age, preoperative opioid use, prescribed opioids at discharge, and hospital length of stay. The increased postoperative day 1 opioid use may be reflective of poorly controlled acute postoperative pain, a known risk factor for developing persistent pain after various surgical procedures.[19–22] For this reason, literature strongly supports the use of multimodal analgesia, early counseling, peripheral nerve blocks as well as neuraxial anesthesia as strategies to minimize the transition from acute to prolonged opioid requirement after surgery.[23–25]

Similar to our findings, preoperative opioid use is also consistently reported to increase the risk of chronic opioid use after surgery.[7–10] Goesling *et al* identified that taking opioids preoperatively, an average daily dose of greater than 60 mg oral MEQs was independently associated with persistent opioid use post lower extremity arthroplasty.[26] Opioid prescription during or after surgery may trigger long-term use in opioid naïve patients.[26–29] Patients who were prescribed greater quantities of opioids at discharge were more likely to request refills for opioids postoperatively. Hernandez *et al* found that the rate of refills did not vary significantly between patients with smaller versus larger opioid prescription,[30] and refills were often prescribed by providers other than the surgeon postoperatively.[31] Excess opioid prescription may also pose the risk of divergence and subsequent abuse.

While we did not identify sex as an important predictive risk factor, age was highly predictive of increased risk. The literature on sex and age as risk factors is variable with some studies finding younger age and female sex associated with persistent opioid need post arthroplasty[32–34] and others suggesting that male sex and older age increase risk of prolonged opioid use.[8–10 35] Similarly, we found that increased body mass index was an important feature predictive of persistent opioid use after lower extremity arthroplasty. This can be secondary to limitation to the progression of inpatient rehabilitation postoperatively[36] or pharmacokinetics, however, outcomes and complications in obese patients are comparable to non-obese patients.[37–39]

By leveraging patient datasets from the EMR, machine learning may be used to offer valuable clinical insight, in this case the prediction of persistent postoperative opioid use after arthroplasty. Such models should then be integrated into a clinical decision support system in the EMR to alert healthcare providers. This predictive model can serve as a foundation for a multidisciplinary transitional pain program, which supports longitudinal care from outpatient postoperative follow-up and long-term analgesia interventions—such as cryoanalgesia or percutaneous neuromodulation—to potentially reduce the likelihood of chronic opioid use.[40] However, in addition, more studies are needed to validate the efficacy of transitional pain clinics, cryoanalgesia, or peripheral neuromodulation on reducing incidence of persistent opioid use. Furthermore, these types of predictive models can also help identify potential subjects into clinical trials designed to enroll high-risk patients only. With the rising interest in early intervention of persistent postsurgical pain, and in anticipation of the emergence of multidisciplinary transitional pain clinics across the country, accurate and reliable predictive analytics technology of at-risk patients becomes cornerstone to this practice of precision medicine.

There are several limitations to our study. Importantly, this is a retrospective study and thus, collection and accuracy of the data was only as good as it was recorded. Therefore, we may have missed some patients that did indeed require persistent opioids due to lack of information in the charts on our review. Future studies would need to develop models from prospectively collected data to ensure accuracy of the features and outcomes. Furthermore, while SMOTE is effective at generating synthetic data to reduce class imbalance,

SMOTE does not take into consideration that neighbors may be from other classes, which can increase noise at the boundary of classes. SMOTE can also be problematic for high dimensional data; however, if variables can be reduced, the bias introduced by the k-NN process will be eliminated.[41] Finally, our predictive models would need to be externally validated in separate datasets to assess its generalizability in this population.

Accurate predictive modeling can provide perioperative physicians with clinical insight to the most vulnerable patient population. Integration of risk factors into an evidence-based perioperative screening tool may allow for early identification of at-risk patients, thus allowing for early intervention by targeted patient-centered systematic approach via a transitional pain program. This approach will achieve the fine balance of addressing acute postoperative pain management, while minimizing the risk of persistent postoperative opioid need.

Contributors RG, BH, SS and ETS were involved in study design. RG, RSP and ETS collected the data. RG, BH and SS were involved in the statistical analysis. RG, BH, RSP, SS, IC and KF were involved in the interpretation of results. RG, BH, RSP, SS, IC, KF and ETS were involved in the preparation and finalization of the manuscript. RG serves as the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This retrospective study was approved by the University of California San Diego's Human Research Protections Program for the collection of data from our electronic medical record system (EMR).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Engy T Said http://orcid.org/0000-0002-7897-1670

REFERENCES

1 CDC, National Center for Health Statistics. *Wide-ranging online data for epidemiologic research (wonder)*. In: Atlanta, GA, 2020.
2 Lawal OD, Gold J, Murthy A, *et al*. Rate and risk factors associated with prolonged opioid use after surgery: a systematic review and meta-analysis. *JAMA Netw Open* 2020;3:e207367.
3 Maradit Kremers H, Larson DR, Crowson CS, *et al*. Prevalence of total hip and knee replacement in the United States. *J Bone Joint Surg Am* 2015;97:1386–97.
4 Beswick AD, Wylde V, Gooberman-Hill R, *et al*. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;2:e000435.
5 Thomazeau J, Rouquette A, Martinez V, *et al*. Predictive factors of chronic post-surgical pain at 6 months following knee replacement: influence of postoperative pain trajectory and genetics. *Pain Physician* 2016;19:E729–41.
6 Lindberg MF, Miaskowski C, Rustøen T, *et al*. Preoperative risk factors associated with chronic pain profiles following total knee arthroplasty. *Eur J Pain* 2021;25:680–92.
7 Goesling J, Moser SE, Zaidi B, *et al*. Trends and predictors of opioid use after total knee and total hip arthroplasty. *Pain* 2016;157:1265.
8 Inacio MCS, Hansen C, Pratt NL, *et al*. Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: a retrospective cohort study. *BMJ Open* 2016;6:e010664.
9 Bedard NA, Pugely AJ, Dowdle SB, *et al*. Opioid use following total hip arthroplasty: trends and risk factors for prolonged use. *J Arthroplasty* 2017;32:3675–9.
10 Bedard NA, Pugely AJ, Westermann RW, *et al*. Opioid use after total knee arthroplasty: trends and risk factors for prolonged use. *J Arthroplasty* 2017;32:2390–4.
11 Cook DJ, Kaskovich SW, Pirkle SC, *et al*. Benchmarks of duration and magnitude of opioid consumption after total hip and knee arthroplasty: a database analysis of 69,368 patients. *J Arthroplasty* 2019;34:638–44.
12 Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. *Journal of Machine Learning Research* 2011;12:2825–30.
13 Chawla NV, Bowyer KW, Hall LO, *et al*. Smote: synthetic minority over-sampling technique. *Jair* 2002;16:321–57.
14 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
15 Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297–302.
16 Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 1948;5:1–34.
17 Fujiwara K, Huang Y, Hori K, *et al*. Over- and under-sampling approach for extremely imbalanced and small minority data problem in health record analysis. *Front Public Health* 2020;8:178.
18 Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* 2013;14:106.
19 Hsia H-L, Takemoto S, van de Ven T, *et al*. Acute pain is associated with chronic opioid use after total knee arthroplasty. *Reg Anesth Pain Med* 2018;43:1–711.
20 van Helmond N, Timmerman H, van Dasselaar NT, *et al*. High body mass index is a potential risk factor for persistent postoperative pain after breast cancer treatment. *Pain Physician* 2017;20:E661–71.
21 Bauer GA, Burgers PM. The yeast analog of mammalian cyclin/proliferating-cell nuclear antigen interacts with mammalian DNA polymerase delta. *Proc Natl Acad Sci U S A* 1988;85:7506–10.
22 Kehlet H, Jensen TS, Woolf CJ. Persistent postsurgical pain: risk factors and prevention. *Lancet* 2006;367:1618–25.
23 Hah JM, Bateman BT, Ratliff J, *et al*. Chronic opioid use after surgery: implications for perioperative management in the face of the opioid epidemic. *Anesth Analg* 2017;125:1733–40.
24 Richman JM, Liu SS, Courpas G, *et al*. Does continuous peripheral nerve block provide superior pain control to opioids? A meta-analysis. *Anesth Analg* 2006;102:248–57.
25 Chitnis SS, Tang R, Mariano ER. The role of regional analgesia in personalized postoperative pain management. *Korean J Anesthesiol* 2020;73:363–71.
26 Goesling J, Moser SE, Zaidi B, *et al*. Trends and predictors of opioid use after total knee and total hip arthroplasty. *Pain* 2016;157:1259–65.
27 Calcaterra SL, Yamashita TE, Min S-J, *et al*. Opioid prescribing at hospital discharge contributes to chronic opioid use. *J Gen Intern Med* 2016;31:478–85.
28 Sun EC, Darnall BD, Baker LC, *et al*. Incidence of and risk factors for chronic opioid use among opioid-naive patients in the postoperative period. *JAMA Intern Med* 2016;176:1286–93.
29 Franklin PD, Karbassi JA, Li W, *et al*. Reduction in narcotic use after primary total knee arthroplasty and association with patient pain relief and satisfaction. *J Arthroplasty* 2010;25:12–16.
30 Hernandez NM, Parry JA, Taunton MJ. Patients at risk: large opioid prescriptions after total knee arthroplasty. *J Arthroplasty* 2017;32:2395–8.
31 Zarling BJ, Yokhana SS, Herzog DT, *et al*. Preoperative and postoperative opiate use by the arthroplasty patient. *J Arthroplasty* 2016;31:2081–4.
32 Singh JA, Lewallen D. Predictors of pain and use of pain medications following primary total hip arthroplasty (tha): 5,707 thas at 2-years and 3,289 thas at 5-years. *BMC Musculoskelet Disord* 2010;11:90.
33 Singh JA, Lewallen DG. Predictors of use of pain medications for persistent knee pain after primary total knee arthroplasty: a cohort study using an institutional joint registry. *Arthritis Res Ther* 2012;14:R248.
34 Singh JA, Gabriel S, Lewallen D. The impact of gender, age, and preoperative pain severity on pain after tka. *Clin Orthop Relat Res* 2008;466:2717–23.
35 Inacio MCS, Pratt NL, Roughead EE, *et al*. Opioid use after total hip arthroplasty surgery is associated with revision surgery. *BMC Musculoskelet Disord* 2016;17:122.
36 Liao C-D, Huang Y-C, Chiu Y-S, *et al*. Effect of body mass index on knee function outcomes following continuous passive motion in patients with osteoarthritis after total knee replacement: a retrospective study. *Physiotherapy* 2017;103:266–75.
37 Agarwala S, Jadia C, Vijayvargiya M. Is obesity a contra-indication for a successful total knee arthroplasty? *J Clin Orthop Trauma* 2020;11:136–9.
38 O'Neill SC, Butler JS, Daly A, *et al*. Effect of body mass index on functional outcome in primary total knee arthroplasty - a single institution analysis of 2180 primary total knee replacements. *World J Orthop* 2016;7:664–9.
39 Boyce L, Prasad A, Barrett M, *et al*. The outcomes of total knee arthroplasty in morbidly obese patients: a systematic review of the literature. *Arch Orthop Trauma Surg* 2019;139:553–60.
40 Katz J, Weinrib A, Fashler SR, *et al*. The Toronto General Hospital transitional pain service: development and implementation of a multidisciplinary program to prevent chronic postsurgical pain. *J Pain Res* 2015;8:695–702.
41 Blagus R, Lusa L. Smote for high-dimensional class-imbalanced data. *BMC Bioinformatics* 2013;14:106.