Pain Medicine 2011; 12: 1566–1575
Wiley Periodicals, Inc.

# Use of Machine Learning Theory to Predict the Need for Femoral Nerve Block Following ACL Repair

**Patrick Tighe, MD,* Sarah Laduzenski, MD,* David Edwards, MD, PhD,* Neal Ellis, MD,* Andre P. Boezaart, MD, PhD,*† and Haldun Aygtug, PhD‡**

Departments of *Anesthesiology and

†Orthopedic Surgery and Rehabilitation, University of Florida College of Medicine, Florida;

‡Information Systems and Operations Department, Warrington College of Business, University of Florida, Gainesville, Florida, USA

*Reprint requests to:* Patrick Tighe, MD, University of Florida College of Medicine, PO Box 100254, Gainesville, FL 32610-0254, USA. Tel: 352-273-7844; Fax: 352-392-7029; E-mail: ptighe@anest.ufl.edu.

Conflict of interest: Dr. A. P. Boezaart invented the stimulating catheter and is receiving royalties from Arrow International (Teleflex Medical) for the sales of StimuCath™.

Presentations: None.

Funding: The funding for this study was from departmental funds.

## Abstract

**Objective. We report on a classification approach using machine learning (ML) algorithms for prediction of postoperative femoral nerve block (FNB) requirement following anterior cruciate ligament (ACL) reconstruction.**

**Background. FNBs are commonly performed for ACL reconstruction to control postoperative pain. Ideally, anesthesiologists would target preoperative FNB only to ACL reconstruction patients expected to experience severe postoperative pain. Perioperative factors associated with postoperative FNB placement following ACL reconstruction remain unclear, may differ among separate surgical facilities, and render such predictions difficult.**

**Methods. We conducted a chart review of 349 patients who underwent ACL reconstruction at a single outpatient surgical facility. Standard perioperative data commonly available during routine preoperative examination were recorded. ML classifiers based on logistic regression, BayesNet, multilayer perceptron, support vector machine, and alternating decision tree (ADTree) algorithms were then developed to predict which patients would require postoperative FNB.**

**Results. Each of the ML algorithms outperformed traditional logistic regression using a very limited data set as measured by the area under the receiver operating curve, with ADTree achieving the highest score of 0.7 in the cross-validated sample. Logistic regression outperformed all other classifiers with regard to kappa statistics and percent correctly classified. All models were prone to overfitting in comparisons of training vs cross-validated samples.**

**Conclusion. ML classifiers may offer improved predictive capabilities when analyzing medical data sets compared with traditional statistical methodologies in predicting severe postoperative pain requiring peripheral nerve block.**

**Key Words. Machine Learning Theory; Femoral Nerve Block; Anterior Cruciate Ligament; Postoperative Pain; Opioids**

## Introduction

Machine learning (ML) is a mathematical algorithm-driven system of classifying large amounts of data into useful information. These models, called "classifiers," take input (i.e., patient symptoms, medications, etc.), process that input, and predict an outcome (i.e., presence or absence of a diagnosis, need for postoperative femoral nerve block [FNB]) based on the underlying patterns. When these patterns are constructed to the event, ML models have been shown to offer expanded prediction capabilities over classical statistical methodologies [1,2]. Modern applications of ML include the complex and data-intensive analysis of handwriting recognition [3], automated voice recognition [4], fraud detection [5], email spam filtering [6], targeted marketing [7], and, most recently, crime fighting.[1]

Preoperative prediction of severe postoperative pain remains an imperfect art, which is notoriously inaccurate if done by doctors [8]. We chose to apply ML to the application of FNB for anterior cruciate ligament (ACL) repair because this is one of the most common ambulatory surgical procedures performed in the United States, with nearly 100,000 projected ACL ruptures annually within the United States alone [9]. Concerns on severe postoperative pain have led many anesthesiologists to routinely place FNBs prior to induction of anesthesia for those patients undergoing ACL repair [10]. However, at some institutions, the associated risks of FNB, including nerve injury, vascular injury, and local anesthetic toxicity, have led to a more conservative approach, with postoperative FNB placed on an as-needed basis [11]. Postoperative FNB following high-dose opioids is suboptimal because the removal of nociceptive input by FNB often results in subsequent sedation and prolonged stay in the postanesthesia care unit. Ideally, FNB would be prospectively targeted in the preoperative setting toward those patients predicted to experience severe postoperative pain to warrant FNB. The challenge, therefore, is to identify the patients who will require a postoperative FNB preoperatively and eliminate the need for high-dose opioid use.

In this study of a retrospective sample, we applied ML classifiers to predict which patients would receive a FNB following ACL repair using only information routinely available prior to the start of surgery. To our knowledge, this is the first time that ML classifiers have been applied to the perioperative prediction of pain. We hypothesized that commonly used ML algorithms, namely support vector machines (SVMs), naive Bayesian analysis, multilayer perceptron, and decision trees [12], would perform comparably with a traditional logistic regression-based analysis. Finally, we examined how these algorithms performed in both the model-training and the model-validation stages of ML classifier development.

## Methods

### *Outcomes*

The primary outcome of this study examined the predictive performance of multiple ML classifiers against a stratified 10-fold cross-validated sample of subjects. The binary classification was the placement, or not, of a postoperative FNB following ACL repair. Secondary outcomes examined the differences between testing and training phases of model development.

### *Description of Subjects*

This study was approved by the Institutional Review Board at the University of Florida. We reviewed the perioperative records of 349 patients who underwent ACL reconstruction at the University of Florida's Florida Surgical Center. All surgeries were performed between August 2007 and August 2009. Subjects received either an open or arthroscopic ACL repair using an allograft or autograft technique harvested from the patellar or hamstring tendon. All procedures were conducted under general anesthesia. Both primary repairs and revisions were included, as were those subjects who received an ACL repair in conjunction with other surgical procedures of the knee. Those subjects who received a preoperative nerve block were excluded from the study sample. The data did not reflect the practice of any single anesthesiologist, nor of an academic subspecialty group, due to the nature of physician scheduling at our institution. FNBs were placed following ACL repair only if the patient continued to report severe pain following multimodal analgesic administration in the recovery room. Multimodal analgesics generally included the use of intraoperative ketamine (10–20 mg i.v.), ketorolac (30 mg), oxycodone (5–10 mg), and intravenous hydromorphone administered in 0.5 mg increments until the patient's pain was either adequately relieved or had required at least 2 mg of hydromorphone within 1 hour of arrival in the recovery room without adequate relief from pain.

### *Input Variables*

Data collected included age, gender, preoperative numerical rating score of pain (0–10), body mass index, and use of alcohol, tobacco, or illicit drugs. The preoperative use of opioids, nonsteroidal anti-inflammatory drugs (NSAIDs), gabapentin, or anxiolytics was recorded as a preoperative patient factor. Recorded surgical characteristics included autograft vs allograft, open vs arthroscopic approach, and total thigh tourniquet time. Recorded anesthetic details included the perioperative administration of gabapentin or pregabalin, NSAIDs, acetominophen, ketamine, or inhalational anesthetic agent as dichotomous inputs. Although anesthetic agent administration was recorded as an intraoperative event, it was included as a preoperative predictor as the decision of which agent to use is commonly made in the preoperative environment.

### *Theory of ML Classification*

The overall mechanism for creating a ML classifier is an iterative process, typically divided into two phases. In the first phase, called the training phase, a set of data with known outcomes is used to train the classifier. The classifier's performance may be overly optimistic in this training phase because the developing model is "fitted" to the same data set with which it was created; this phenomenon is referred to as "overfitting." In the second phase, called the testing phase, the classifier's performance is measured by using it to make predictions on a previously unseen data set (also known as out-of-sample testing). Because the classifier is applying lessons learned from the training data set to a new set of instances in need of classification, the ML classifier's performance may decrease to reflect a more realistic situation. In some respects, this testing phase is similar to the validation phase of any clinical prediction tool. A "good classifier" is

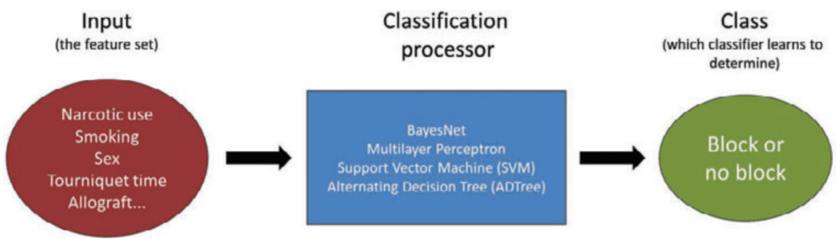


**Figure 1** Overview of machine learning (ML) classification implementation. Input variables are defined for each instance and then entered into one or more ML classifiers. The classifiers then render predictions on the class to which the instance should belong based on the training of the classifier.

one that performs well during both training and testing phases. Although earlier work on classification algorithms focused on logistic regression and neural networks, recent efforts have expanded to include decision trees, SVMs, and Bayesian analysis. A broad overview of this process is given in Figure 1.

A variety of software products have been developed to assist with ML classifier application, yet each approaches the problem of classification using a similar approach. Examples include RapidMiner (Rapid-I, Dortmund, Germany), Weka (University of Waikato, New Zealand), and Enterprise Miner (SAS Institute, Cary, NC, USA). In this study, all ML algorithms were run using WEKA 3.6.2 [12]. Each implementation first requires the user to import data sets either as separate training and testing (or hold-out) sets or as a single set that can be partitioned into training and testing phases. The user then selects a cohort of ML classifiers to develop using the training data. Once refined, the ML classifiers then classify the testing (hold-out) data, which have known outcomes but were not used for classifier development. This allows objective scoring of ML classifier performance. We used the above methodology throughout this study. We used 10-fold cross-validation for testing ML classifiers against the hold-out data [13].

Due to the pilot nature of this project, we assembled a team of mathematicians, anesthesiologists, and administrators to assist with problem formulation, classifier selection and tuning, and interpretation of results. Because ML classification is heavily related to the discipline of data mining, there is ongoing debate concerning the role of traditional biostatistics vs applied data mining [14]. This is similar to the debate between Bayesian and classical statistics [15].

### *Description of Classifiers*

Our study compared the results across four classifiers: BayesNet, multilayer perceptron, SVM, and alternating decision tree (ADTree). Additionally, we included a logistic regression analysis as a comparator. The simple logistic regression implemented in Weka used no errors on probabilities, heuristic stop at 50, maximum boosting iterations of 500, with cross-validation of logit boost iterations [16].

The BayesNet classifier approaches the problem of classification as one of the conditional probabilities. It constructs a directed acyclic graph or a series of nodes connected by directional links that do not form loops. Each node represents a variable, and the linkages the conditional probabilities among the nodes (Figure 2). Our classifier used the K2 search algorithm, which systematically tries to add "parent" nodes to each node to improve the resulting graph, as well as a Markov blanket classifier that permitted grouping of parent and child nodes. The BayesNet classifier was run using the SimpleEstimator with an alpha of 0.5, a maximum of one parent node, and random ordering of network nodes.[2]

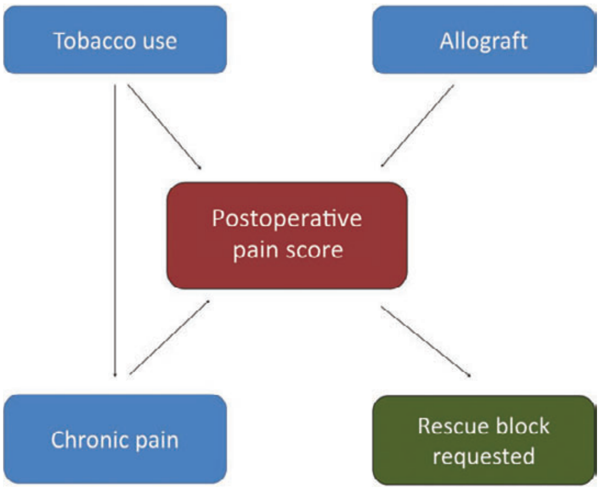


**Figure 2** Schematic representation of a BayesNet network. The BayesNet algorithm constructs a directed acyclic graph or a series of nodes connected by directional links that do not form loops. Each node represents a variable, and the linkages between nodes the conditional probabilities among the nodes.

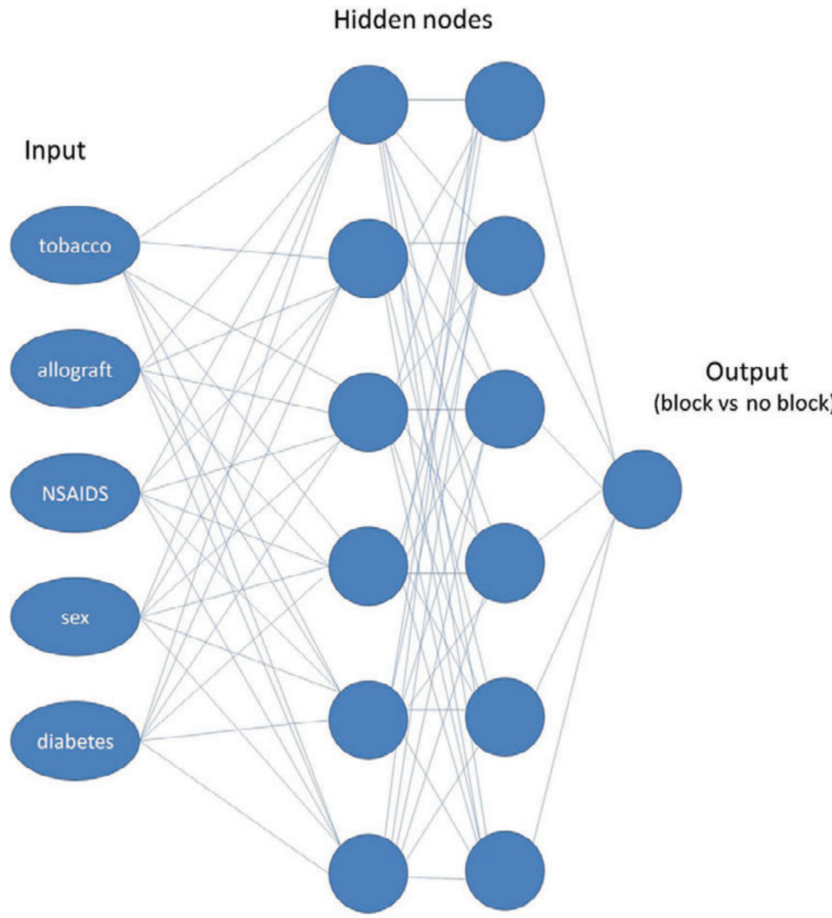


**Figure 3** Schematic of a prototypical multilayer perceptron. Following inputs, data progress to hidden layers using different functions, which can be either logic- or function-based. Each node functions as a neuron, passing information to the next node only if the information reaches a certain threshold. The difference between the eventual outcome and the known outcome, or error, leads the perceptron to backpropagate this information, permitting the perceptron to adjust its nodes in an attempt to reduce the final error. This process is repeated iteratively.

Multilayer perceptrons are a type of neural network. First, an input layer of nodes is constructed, followed by multiple levels of hidden layers, and finally an output layer. Information is passed from one layer of nodes to the next, but as with neurons, nodes will only transmit the information to the next node if the information reaches a certain threshold of importance. By comparing the actual outputs with the classification goal, the thresholds required for node activation are iteratively updated via backpropagation until a final optimization is reached (Figure 3) [17]. In this case, the multilayer perceptron was set to autobuild with two hidden layers, learning rate of 0.3 and momentum of 0.2, nominal to binary filter off, normalized attributes and normalized numeric classes, seed of 0, and training series of 1,000,000 with validation size of 10% and validation threshold of 20 [13,18].

SVMs are a function-based classifier, which first plot all instances on a coordinate system and then derive a function to maximally separate the instances according to the target classification scheme. The complexity of this function is influenced by the complexity value, and the type of function by the selected kernel (Figure 4). The support vector classifier incorporating a logistic model used a complexity value of 1, the round-off error or epsilon of $1 \times 10^{-12}$, normalized training data filter, the Puk kernel based on the Pearson VII distribution with omega and sigma both set to 1, random seed of 1, and tolerance parameter of 0.001 [19,20].

The ADTree is a type of decision tree. With the ADTree, the decision tree establishes prediction nodes each consisting of single numbers, and then decision nodes that predicate a decision on the prediction node based on the input variables. Each instance is then applied to the tree, and the prediction nodes through which the instance passes are summed to render a score. The magnitude of this score dictates the classification. The ADTree uses boosting or the combination of multiple "weak" algorithm iterations to create a stronger classifier (Figure 5). The ADTree classifier used 10 boosting iterations with random seed of 0 and default expansion of all paths in designing the decision tree.[3]

### *Statistical Analyses*

General statistical comparisons for univariate analysis were completed using JMP 8.0.2 (SAS Institute). The *t*-test and chi-square analysis were used for univariate comparisons as appropriate. A two-sided Fisher's exact result was substituted for routine chi-square analyses for

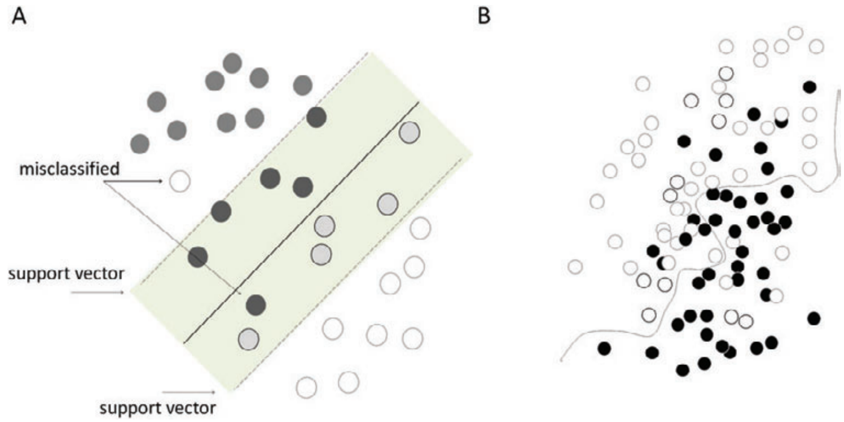


**Figure 4** Schematic representation of a support vector machine (SVM). Circles of different colors represent instances belonging to different classes. The SVM attempts to design a function describing a line (panel A [high-dimension space]) or a plane (panel B [kernel-based hyperplane mapping in the input space]), which can separate instances according to class. This plane can extend into very high-dimensional space as necessary, but is visualized here in two dimensions (panel B).

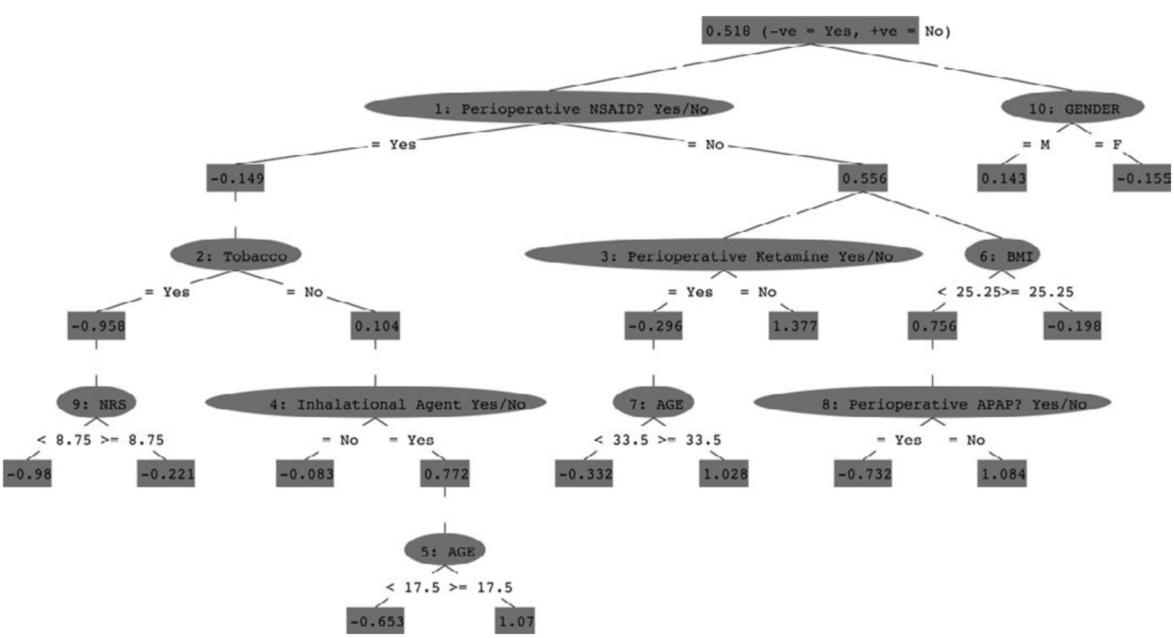


**Figure 5** Schematic representation of an alternating decision tree (ADTree). With the ADTree, the algorithm establishes prediction nodes each consisting of single numbers and then decision nodes that predicate a decision on the prediction node based on the input variables. Each instance is then applied to the entire tree, and the prediction nodes through which the instance passes are summed to render a score. The magnitude of this score dictates the classification. The tree is "boosted" in that a number of trees are created, the best scores from each tree combined into a single classification. NRS = numerical rating scale; NSAID = nonsteroidal anti-inflammatory drug; BMI = body mass index; APAP = acetaminophen.

the comparisons involving illicit drug use, home anxiolytic use, open surgical repair, and perioperative acetominophen administration due to expected cell counts of less than five in a $2 \times 2$ table.

We reported the results of both training and testing phases of ML development to demonstrate the capabilities and risk of overfitting of ML classifiers. Multiple metrics were incorporated into the evaluation of each classifier. The overall percent of correctly classified instances reflects a simple evaluation of a classifier, the same as evaluation by the area under a receiver operating curve (ROC). Because a classifier relying on random selection of instances will frequently classify some instances correctly, we used the kappa statistics to control for those instances that may have been correctly classified only by chance [21]. We also evaluated the accuracy of each classifier by its *F*-measure, which represents the harmonic mean between precision and recall [22]. For the test phase, all models used stratified 10-fold cross-validation [13].

### Results

Figure 6 describes the methodology for data aggregation from the perioperative chart. Overall, 91 out of 349 patients (26%) required a postoperative nerve block. Baseline characteristics of the study sample based on the need for a postoperative FNB are shown in Table 1. Based on the univariate analysis, female gender and tobacco use were increased in those needing rescue FNB. There was no difference in surgical factors (autograft vs allograft, open repair, tourniquet time) between groups. Perioperative NSAID, ketamine, gabapentin, or pregabalin were all used more frequently in those receiving a rescue FNB, although inhalational agents were used less often.

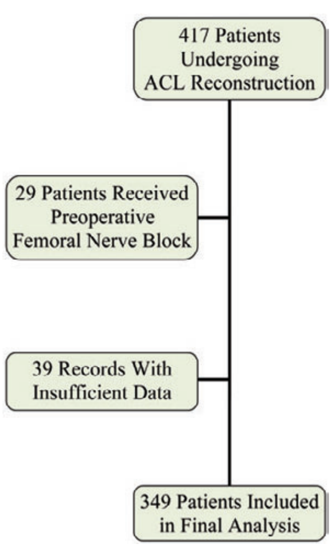


**Figure 6** Review of data aggregation from medical records. Records with insufficient data included those missing information on outcomes, thus rendering them ineligible for classification in either machine learning classification training or testing stages. ACL = anterior cruciate ligament.

Table 2 compares the training and cross-validation results for each of the models studied using data available to the anesthesiologist in the preoperative setting. ROC area was greatest in the cross-validated data set using the ADTree classifier, while SVM had the highest kappa value. Logistic regression outperformed other classifiers in the percent correctly classified as well as the *F*-measure.

The logistic regression algorithm used by WEKA uses a boosting option, and we permitted up to 500 boosting iterations. Notably, boosting is not used in all implementations of logistic regression. When minimized to a single-boosting iteration, the returned ROC area was only 0.559, a notable decrease from the 0.645 given using the default settings of the simple logistic implementation of logistic regression.

### Discussion

Our results suggest associations between perioperative factors and the need for postoperative FNB placement following ACL reconstruction. Although logistic regression demonstrated appropriate classification ability, other ML classifiers offered comparable, or better, classification performance during both training and cross-validation. This similarity may underestimate the capabilities of the ML classifiers because our data set did not include data with a high ratio of attributes to instances, where ML may offer significant improvements over simple logistic regression [2]. For instance, an elevated attribute-to-instance ratio would include a data set of subjects whereby 100 subjects were included, yet each subject had over 50 variables such as age, gender, and zip code. Contrariwise, traditional classifiers such as logistic regression can manage only one variable per 10 outcome events [23].

Our analysis used relatively default configurations for each classifier; no attempt was made to significantly optimize any one model. Although optimization of a classifier would have likely improved classification performance, we hoped to demonstrate the efficacy of “out of the box” classifiers vs those that would be incrementally tuned. Additionally, a focus on multiple performance metrics would have lead to multiple versions of each classifier tuned against a specific outcome rather than addressing the original outcomes. Future investigations may wish to compare baseline algorithm performance against the performance of systematically tuned algorithms to determine the extent of improvement possible.

Traditionally, statistical tests have been geared toward testing of a single hypothesis, and data mining as implemented by ML classifiers focused on the search through multiple potential hypotheses [1,18]. The ML classifiers included in this study encompass a wide variety of

**Table 1** Univariate analysis of patient, surgical, and anesthetic factors associated with postoperative femoral nerve block

| | | Rescue block? | | |
|---|---|---|---|---|
| | | No | Yes | *P* value |
| Patient factors | Age | 26.5 | 26.5 | 0.9994 |
| | Female | 33% | 48% | 0.012† |
| | NRS | 1.37 | 1.63 | 0.51 |
| | BMI | 26 | 27.6 | 0.0635 |
| | EtOH (responding yes) | 6.98% | 6.59% | 0.9 |
| | Tobacco (responding yes) | 4.65% | 19.8% | <0.0001† |
| | Illicit drug (responding yes) | 1.55% | 3.3% | 0.3829‡ |
| | Home opioid use | 4.26% | 7.69% | 0.22 |
| | Home NSAID use | 19.38% | 24.18% | 0.337 |
| | Home gabapentin/pregabalin | 12.02% | 15.38% | 0.4172 |
| | Home anxiolytic use | 1.16% | 1.11% | 1‡ |
| Surgical factors | Autograft repair | 32.95% | 27.47% | 0.3299 |
| | Allograft repair | 67.05% | 71.43% | 0.4382 |
| | Open repair | 3.88% | 4.40% | 0.764‡ |
| | Tourniquet time (minutes) | 48.3 | 45.9 | 0.2321 |
| Anesthetic factors | Perioperative NSAID | 65.5% | 89% | <0.0001† |
| | Perioperative APAP | 4.26% | 6.59% | 0.3988‡ |
| | Perioperative ketamine | 75.19% | 87.91% | 0.0049† |
| | Perioperative gabapentin/pregabalin | 25.58% | 42.86% | 0.0024† |
| | Inhalational agent | 17.83% | 8.79% | 0.0313† |

† Significant at the 0.05 level.
‡ Test for significance using Fisher's exact test.
APAP = acetaminophen; BMI = body mass index; EtOH = ethanol; NSAID = nonsteroidal anti-inflammatory drug; NRS = numerical rating scale.

mathematical underpinnings. Despite their differing approaches, each classifier returned similar results on the cross-validated data. Each classifier, but especially the SVM, "over-fit" the training-set data. This reflects the remarkable adaptability of the SVM to identify a very complex function, which defines a surface separating each instance according to the outcome in question. In other words, on a Cartesian coordinate system, one might imagine that all instances resulting in a postoperative block lie above this plane, and all instances without a block below the plane. As with any overfit model, the SVM has adapted itself so perfectly to the training set, including outliers, that it becomes too rigid for future data, which may not perfectly follow the identified function. One example would be the athlete who simply wants a nerve block after surgery due to the positive reviews offered by her teammate. No volume of biopsychosocial indicators would likely predict such a circumstance, yet such an

**Table 2** Training and cross-validation results for classification algorithms, preoperative data inputs only

| Stage | | BayesNet | Multilayer perceptron | SVM | ADTree | Simple logistic regression |
|---|---|---|---|---|---|---|
| Training | ROC area | 0.737 | 0.778 | 0.996 | 0.772 | 0.657 |
| | Kappa statistics | 0.295 | 0.320 | 0.902 | 0.301 | 0.228 |
| | Percent correctly classified | 77.4% | 79.7% | 96.3% | 79.1% | 77.7% |
| | *F*-measure | 0.744 | 0.754 | 0.962 | 0.747 | 0.72 |
| 10-fold cross-validation | ROC area | 0.691 | 0.688 | 0.671 | 0.7 | 0.645 |
| | Kappa statistics | 0.2169 | 0.1703 | 0.242 | 0.179 | 0.228 |
| | Percent correctly classified | 75.1% | 74.5% | 65.9% | 74.2% | 77.7% |
| | *F*-measure | 0.716 | 0.7 | 0.677 | 0.702 | 0.72 |

ADT = alternating decision tree; ROC = receiver operating curve; SVM = support vector machine.

instance in a training set might doom an SVM plane into extreme contortions to satisfy this event.

Aside from the above exception, the need for a postoperative FNB frequently indicates the presence of severe postoperative pain. The level of pain and opioid requirement may not be well predicted by the type or complexity of the ACL repair [24]. Werner et al. recently conducted a meta-analysis of studies attempting to predict postoperative pain based on perioperative predictors [8]. They concluded that up to 54% of the variance in postoperative pain may be predicted by preoperative evaluations of sensory testing, nociception, anxiety, and catastrophic thinking. Although quantitative sensory testing may offer specific advantages in prediction of postoperative pain [25–27], their implementation in routine preoperative assessment is far from universal. Our study focused on only those variables routinely captured in the perioperative setting. Expansion of predictors into the intraoperative time frame may improve ML classifier performance, yet render predictions too late for preoperative nerve block. Still, a prediction rendered at the end of a surgical procedure may still assist the clinician in deciding to place a FNB immediately following emergence from anesthesia vs a trial of systemic analgesics within the recovery room.

Our economical use of existing data vs exploration of additional biological predictors, each requiring discrete investigation, highlights an important methodological difference in our approach to the problem of postoperative pain prediction. We were able to examine improvements in clinical decision support in a manner that required no additional testing of patients. Clinicians rarely make decisions based on single or even multiple individual characteristics. More importantly, the specific evidence behind such predictors, especially when studied in carefully controlled environments such as randomized controlled clinical trials (RCT)s, lose certainty as they are applied to foreign situations not encountered in the original RCTs. Clinicians instead combine their prior experience with simultaneous input from dozens of specific predictors, some of which may be considered “evidence-based.” Critically, ML classifiers still rely on relevant characteristics to render their classifications. Rather than offer a substitute for biopsychosocial research platforms, we envision ML classifiers as a complement, assisting with the translation of such findings into improvements in clinical care.

We envision a system that will automatically pull a test data set, classify this test data against ML classifiers designed using historical training data, render predictions, and then report such predictions to the pertinent anesthesiologists via a nightly email. The next-day experience would then be entered into the historical training set to improve the ML classifier’s performance on the next day. Regardless of the problem set, the same two-phase methodology of ML classifier training and refinement, followed by testing and updating, would apply.

Our study suffered some limitations, chiefly its retrospective nature. Only those variables available within the perioperative record were available for entry into the classification algorithms. Also, we believe the name of the surgeon, who was not used as a variable in this study, may have a significant influence on the outcome. We may even speculate that using surgeon name as a variable could make the predictions substantially more accurate. Furthermore, it is possible that the predictive ability would be greatly improved by inclusion of formalized quantitative sensory and psychosocial evaluations. Although many additional variables conceivably could have been included, the restricted sample size prohibited the rationale inclusion of many factors considered to be minimal in impact or inappropriately partitioned between groups. This particular limitation is more specific to logistic regression than ML classifiers, which may retain discriminatory power even when the number of covariates is greater than the number of data points [1]. Further studies using prospective implementation of ML classifier-based forecasting of postoperative pain are required to fully determine the impact of ML classifiers for improving perioperative patient care.

One perceived weakness to our classification approach is the absence of single-unit predictors. Traditionally, classical statistics has left the reader with a few covariates that were deemed statistically significant following some type of multivariate regression. However, such approaches fail to include these covariates in a conditional framework, or as high-level interactions, as demonstrated in Figure 5.

Perhaps one of the most interesting limitations pertains to the ramifications of clinically oriented ML classification systems. Currently, no standards exist for ML classifier-based analytics, although some entities recognize software-focused training certifications. The complex nature of ML classifier algorithm development and deployment will likely restrict full understanding to those physicians without fellowship training unless medical training is significantly revised to include substantial increases in mathematics and statistics. In the United States, federal incentives to both deploy electronic medical record systems and use them as clinical decision support tools may expedite the development and utilization of clinical ML classification systems.[4]

In conclusion, our study represented a proof-of-concept, demonstrating the performance of ML classifiers against more common classification methodologies involving logistic regression. Our efforts focused on the economic use of available information rather than prospectively studied single-predictor attributes. Regardless of the problem set, the same two-phase methodology of ML classifier training and refinement, followed by testing and updating, would apply. Although our work focused on postoperative nerve block requirements, the same methodologies could be used to predict patient satisfaction with pain control, severe postoperative pain scores, prolonged length-of-hospitalization, postoperative respiratory failure, or any other health care problem requiring complex and highly interactive, or even conditionally related, predictors to establish classifications. Due to the complex nature of perioperative pain, forecasting efforts will likely necessitate novel integration of predictive attributes such as those provided by ML classifiers.

## Notes

1. Heivilin LJ. The First Predictive Policing Symposium: Symposium Report. November 18–20, 2009, Los Angeles CA. Available at: http://www.iaca.net/Articles/First_Predictive_Policing_Symposium_Report_Final.pdf.

2. Bouckaert RR. Bayesian Network Classifiers in Weka for Version 3-5-7. May 12, 2008. Available at: http://www.cs.waikato.ac.nz/~remco/weka.bn.pdf (accessed August 19, 2010).

3. Freund Y, Mason L. The alternating decision tree learning algorithm. In: Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, 124–133, 1999.

4. Centers for Medicare & Medicaid Services. CMS Finalizes Definition of Meaningful Use of Certified Electronic Health Records (EHR) Technology. Available at: https://www.cms.gov/apps/media/press/factsheet.asp?Counter=3794&intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500&srchOpt=0&srchData=&keywordType=All&chkNewsType=6&intPage=&showAll=&pYear=&year=&desc=&cboOrder=date (accessed July 19, 2011).

## References


1 Breiman L. Statistical modeling: The two cultures. Stat Sci 2001;16:199–215.

2 Westreich D, Lessler J, Funk MJ. Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol 2010;63:826–33.

3 Labusch K, Barth E, Martinetz T. Simple method for high-performance digit recognition based on sparse coding. IEEE Trans Neural Netw 2008;19:1985–9.

4 Hamill M, Young V, Boger J, Mihailidis A. Development of an automated speech recognition interface for personal emergency response systems. J Neuroeng Rehabil 2009;6:26.

5 Bolton RJ, Hand DJ. Statistical fraud detection: A review. Stat Sci 2002;17:235–55.

6 Zorkadis V, Karras DA, Panayotou M. Efficient information theoretic strategies for classifier combination, feature extraction and performance evaluation in improving false positives and false negatives for spam e-mail filtering. Neural Netw 2005;18:799–807.

7 Bose I, Mahapatra RK. Business data mining—A machine learning perspective. Inf Manag 2001;39:211–25.

8 Werner MU, Mjöbo HN, Nielsen PR, Rudin A. Prediction of postoperative pain: A systematic review of predictive experimental pain studies. Anesthesiology 2010;112:1494–502.

9 Prodromos CC, Han Y, Rogowski J, Joyce B, Shi K. A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen. Arthroscopy 2007;23:1320–5.

10 Iskandar H, Benard A, Ruel-Raymond J, Cochard G, Manaud B. Femoral block provides superior analgesia compared with intra-articular ropivacaine after anterior cruciate ligament reconstruction. Reg Anesth Pain Med 2003;28:29–32.

11 Mall NA, Wright RW. Femoral nerve block use in anterior cruciate ligament reconstruction surgery. Arthroscopy 2010;26:404–16.

12 Hall M, Frank E, Holmes G, et al. The WEKA data mining software: An update. SIGKDD Explor 2009;11:10–8.

13 Efron B. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983;37:36–48.

14 Breiman L. Statistical modeling: The two cultures. Stat Sci 2011;16:199–231.

15 Lee S-M, Abbott P, Johantgen M. Logistic regression and Bayesian networks to study outcomes using large data sets. Nurs Res 2005;54:133–8.

16 Sumner M, Frank E, Hall M. Speeding up logistic model tree induction. In: Jorge A, Torgo L, Brazdil P, Camacho R, Gamma J, eds. Knowledge Discovery in Databases: PKDD. 2005. Berlin/Heidelberg: Springer; 2005:675–83.

17 Bishop CM. Neural Networks for Pattern Recognition. New York: Oxford University Press; 1995.

18 Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. San Francisco, CA: Morgan Kaufmann; 2005.

19 Üstün B, Melssen WJ, Buydens LMC. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. Chemometr Intell Lab 2006;81:29–40.

20 Yang ZR. Biological applications of support vector machines. Brief Bioinform 2004;5:328–38.

21 Carletta J. Assessing agreement on classification tasks: The kappa statistic. Comput Linguist 1996;22:249–54.

22 Hripcsak G, Rothschild AS. The practice of informatics: Technical brief. Agreement, the F-measure, and reliability in information retrieval. J Am Med Inform Assoc 2005;12:296–8.

23 Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373–9.

24 Beck PR, Nho SJ, Balin J, et al. Postoperative pain management after anterior cruciate ligament reconstruction. J Knee Surg 2004;17:18–23.

25 Granot M, Lowenstein L, Yarnitsky D, Tamir A, Zimmer EZ. Postcesarean section pain prediction by preoperative experimental pain assessment. Anesthesiology 2003;98:1422–6.

26 Lundblad H, Kreicbergs A, Jansson KA. Prediction of persistent pain after total knee replacement for osteoarthritis. J Bone Joint Surg Br 2008;90:166–71.

27 Nielsen PR, Norgaard L, Rasmussen LS, Kehlet H. Prediction of post-operative pain by an electrical pain stimulus. Acta Anaesthesiol Scand 2007;51:582–6.