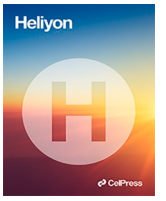
Heliyon 10 (2024) e40602

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Machine learning approach to predict postoperative pain after spinal morphine administration during caesarean delivery

Chin Wen Tan [a,b,*], Juan Zhen Koh [c], Hanwei Jin [c], Nian-Lin Reena Han [d], Shang-Ming Cheng [a,b], Andy Wee An Ta [c], Han Leong Goh [c], Ban Leong Sng [a,b]

[a] Department of Women's Anaesthesia, KK Women's and Children's Hospital, Singapore
[b] Anaesthesiology and Perioperative Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore
[c] Data Analytics and AI, Synapxe Pte Ltd, Singapore
[d] Division of Clinical Support Services, KK Women's and Children's Hospital, Singapore

ARTICLE INFO

Keywords:
Caesarean
Logistic regression
Machine learning
Postoperative pain

ABSTRACT

Background: A major barrier to optimal pain management is the difficulty in predicting and assessing patients at high risk for significant pain across multiple locations within the institution in a timely manner. This is compounded by the fragmented display of clinical information on enterprise clinical platform, which further hinders delay the reviews and hence the increased risk of untreated pain. We evaluated and compared the predictive performance of six modelling techniques in predicting significant pain, defined as the maximum pain score of 3 or more on movement at the 13th to 24th hour after spinal morphine administration during caesarean delivery.

Methods: We retrieved medical records from women who underwent caesarean delivery and received postoperative spinal morphine in a single specialist maternity hospital in Singapore between Aug 2019 and Aug 2022. We extracted 120 clinical variables from the medical records of eligible patients and further selected 23 to improve algorithm accuracies. The dataset was split randomly, with 80 % of patients (n = 5248) used for training the models, and 20 % (n = 1313) reserved for validation.

Results: The study cohort comprised 6561 patients with an incidence of significant postoperative pain of 7.9 %. Ridge regression demonstrated the best performance with both the full (AUC: 0.649) and selected (AUC: 0.719) feature sets. By reducing the number of features, Ridge regression, LASSO, Elastic net, and XGBoost showed similar in AUC (0.704–0.719), sensitivity (0.644–0.695), specificity (0.644–0.705), positive predictive value (0.155–0.179), and negative predictive value (0.949–0.955) in predicting significant postoperative pain. These were attributed to the top three variables, mainly the last recorded postoperative pain score (on movement) before the prediction point, mean and standard deviation of the hourly maximum postoperative pain score (at rest) at 0^{th} to 12th hour.

Conclusions: Future research will aim to refine these models and explore their implementation in clinical settings to enhance real-time pain management and risk stratification for women after caesarean delivery.

* Corresponding author. Department of Women's Anaesthesia, KK Women's and Children's Hospital, 100 Bukit Timah Road, 229899, Singapore.
E-mail address: tan.chin.wen@kkh.com.sg (C.W. Tan).

https://doi.org/10.1016/j.heliyon.2024.e40602
Received 24 July 2024; Received in revised form 13 November 2024; Accepted 20 November 2024
Available online 22 November 2024

1. Introduction

Postoperative pain management after caesarean delivery is crucial for enhancing recovery, breastfeeding, mother-child bonding, and overall satisfaction [1]. Spinal morphine is commonly administered for treating acute post-caesarean pain; yet it does not eliminate the risk of significant acute postoperative pain after caesarean delivery for all patients [2]. It has been reported that experiencing severe acute pain after caesarean delivery may increase the risk of postoperative persistent pain [3]. Defined as surgical site pain lasting beyond 3 months, with other causes of pain excluded, postoperative persistent pain can result in long-term adverse physical and psychological consequences. Persistent pain after caesarean delivery is associated with compromised functional activities, prolonged maternal recovery, and poor quality of life [3]. Given that 21 % of worldwide deliveries are performed via caesarean delivery, this presents a substantial healthcare burden due to increased demand and costs for postoperative support services and visits [4].

One of the main barriers preventing optimal pain management is the inability to predict and provide early assessment and intervention for patients at greater risk of significant pain in a timely manner. At our institution, acute pain service providers review the patients within one day after surgery and regularly document pain scores. Clinical information (vital signs, analgesic medications administered, surgical, and patient demographic data, etc.) can be accessed on a secure enterprise clinical platform available to healthcare personnel. Currently, this information is displayed in different domains within the platform, making it difficult for healthcare personnel to prioritize patient assessments. As clinicians often need to review a large number of patients across multiple locations within the hospital, those with significant pain may not be prioritized for early assessment, which can increase the risk of untreated pain.

Machine learning has been used to develop risk stratification predictive models for postoperative pain, owing to its capability to identify patterns within large datasets with complex interactions. Nair et al. identified Random Forest as the best algorithm to predict opioid requirements among patients after ambulatory surgery [5]. Tighe et al. found that linear regression with Least Absolute Shrinkage and Selection Operator (LASSO) outperformed in predicting postoperative pain for general non-ambulatory surgery [6]. However, these studies have several limitations: i) the models/algorithms were not generalizable and not applicable to caesarean deliveries; and ii) the developed models may not be applicable in Singapore's multi-ethnic population.

To address these issues and improve postoperative pain outcomes, we identified predictors and developed risk stratification predictive models for significant postoperative pain after caesarean delivery with spinal morphine administration. We compared several machine learning algorithms including regularized logistic regression models (ridge regression, LASSO, and Elastic net) and ensemble learning algorithms (Random Forest, XGBoost, and LightGBM).

2. Methods

2.1. Ethics and perioperative management protocols

We conducted a retrospective analysis using deidentified perioperative database ("SingHealth-IHiS Electronic Health Intelligence System (eHIntS)") that recorded patient, surgical, anaesthesia and postoperative data obtained from women aged 18–50 years old who had caesarean delivery with spinal morphine administration for pain management between Aug 2019 and Aug 2022 in KK Women's and Children's Hospital, a Singapore's major public maternity institution. The study received a waiver of consent from the SingHealth Centralised Institutional Review Board ((10 Hospital Boulevard #19-01 SingHealth Tower Singapore 168,582; reference number CIRB 2022/2505; chairperson Dr Steve Yang; approved on Sep 22, 2022). This study adheres to the applicable Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [7].

Perioperative management was administered at the discretion of the attending clinicians in accordance with the hospital's routine standards of care for caesarean delivery. In brief, patients received spinal anaesthesia using 0.5 % bupivacaine with dextrose as the anaesthetic agent. All patients were administered spinal bupivacaine, along with intrathecal 15 μg fentanyl and 100 μg morphine as adjuncts for intraoperative anaesthesia. Testing and monitoring of the spinal block and pain were performed at various time points throughout the surgery. During the postoperative period, patients were discharged from the post-anaesthesia care unit (PACU) to respective ward upon resolution of the spinal block to T4 or below. The postoperative analgesics including paracetamol and mefenamic acid were administered following routine clinical practice, while tramadol was provided upon patient request for additional pain relief with pain score 3 or more.

The primary outcome of significant pain was defined as whether a patient experienced a maximum postoperative numerical rating scale (NRS) pain score on movement of 3 or more (from a scale of 0–10) at 13th to 24th hour after spinal morphine administration. A cut-off of 3 is considered clinically relevant in clinical practice, including our institution, as those who score 3 and above would require additional pain assessment and treatment (e.g., tramadol which is provided in our institution) [8]. If the data were missing in the primary outcome, the records were excluded from the analysis.

2.2. Cohort selection and variables definition

We randomly selected 80 % of the patient data (n = 5248) as the training cohort to develop the risk stratification models for postoperative pain after spinal morphine administration, while the remaining 20 % (n = 1313) were used for model validation. To address class skewness and reduce bias during machine learning, random undersampling of the majority class without replacement was implemented on the training cohort [9].

Independent variables from six categories were retrieved from electronic medical records and computed based on the information

prior to the 13th hour after spinal morphine administration: patient characteristics, surgery information, vital signs, PACU information, medication records, and patients' self-reported pain scores (Fig. 1). The extracted data were coded and deidentified to protect patient privacy and confidentiality. Data reviews were conducted to identify any missing data in the independent variables. For unavailable data, categorical variables were labelled as missing, whereas ordinal and continuous variables were filled using mode and mean, respectively.

Continuous variables were summarized as mean with standard deviation (SD), while medians with interquartile range (IQR) were used as alternative for skewed data distributions. Categorical variables were presented as frequencies with corresponding proportions. Statistical significance was set at $P < 0.01$, and all tests were two-tailed. Statistical analysis and machine learning algorithms were developed in Python (version 3.11.5) using the Pandas (version 1.5.3), NumPy (version 1.26.4), featurewiz (version 0.5.5) and scikit-learn (version 1.2.2) packages. Featuring engineering was conducted to convert the deidentified patient data into clinically relevant variables based on clinician inputs. A total of 120 variables were created and summarized into broad categories (Appendix S1).

2.3. Feature selection

Feature selection optimization is the process of identifying the most relevant input variables that are important to the prediction tasks while minimizing redundancy [10]. Multicollinearity was found from our pairwise Pearson correlation analysis performed on our 120 variables, with 37 pairs of variables having absolute correlation coefficient greater than 0.9. To address this challenge of multicollinearity, we employed the Searching for Uncorrelated List of Variables (SULOV)-recursive method to detect highly correlated pairs of variables, evaluate their importance to the target, and remove the less informative variable of each pair from the feature set [11]. By shortlisting the variables with the highest informational value and the least correlation with one another, we parsed the features to the recursive XGBoost algorithm to identify the most predictive variables in an iterative manner using five training-validation cycles, with the top features collated from varying data subsets. All these operations were performed using the "featurewiz" python package, resulting in the selection of a total of 23 features for further model development (Table 1).

2.4. Predictive model development and validation

We implemented and compared six machine learning models, including the regularized logistic regression models (ridge regression, LASSO, and Elastic net) and the ensemble learning algorithms (Random Forest, XGBoost, and LightGBM) (Table 2).

In the context of logistic regression, increasing the number of features and coefficients may capture more intricate relationships between predictor variables and the dependent variable. However, this also raises the likelihood of overfitting to the training data and modelling random noise [12]. Ridge Regression, LASSO and Elastic net are different regularization techniques used to balance the bias-variance trade-off in machine learning by adding a penalty term to the loss function of the logistic regression [13]. Ridge regression, is particularly useful in situations with a high degree of multicollinearity among predictor variables, as it can shrink the coefficients toward zero [14]. LASSO regression penalizes the absolute values of individual coefficients, allowing some to shrink to zero; hence effectively performing feature selection by removing those features from the model [12]. Elastic net, on the other hand, combines both regularization techniques into a single loss function, making it useful for handling highly correlated features while facilitating feature selection [15].

Random Forest, XGBoost and LightGBM are supervised ensemble learning methods used for classification and regression tasks. The Random Forest model trains multiple decision trees with different bootstrap samples in parallel, aggregating their predictions based on the majority votes of individual trees [16]. This method, known as bagging, helps reduce overfitting during model training. In contrast, XGBoost uses adaptive boosting to sequentially integrate weak classifiers using a gradient descent algorithm. Lastly, LightGBM, known for its relatively fast gradient boosting technique on large datasets, grows trees leaf-wise based on the maximum delta loss [17].

For each model, we applied five-fold cross-validation with a grid search to tune the hyperparameters and reduce overfitting in the training dataset. K-fold cross-validation is a resampling technique that divides data into k partitions, trained on and validated k times, each time using a different subset for validation and the remaining data for training [18]. To evaluate the performance of the models, all six predictive models were assessed by receiver operating characteristics (ROC) analysis with the validation cohort. The overall predictive performances of the models were compared using the area under the curve (AUC) metric, which illustrates the trade-off between the true positive rate and false positive rate. In addition to AUC, sensitivity, specificity, positive predictive value (PPV),

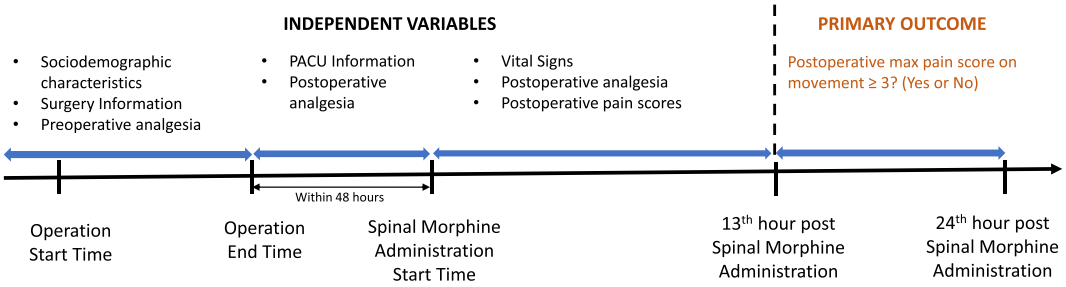


Fig. 1. Timeline view of variables used for the study.

Table 1
The top 23 features selected using SULOV-recursive method for model development.

Patient Characteristics	Age (y)
	Height (cm)
	Body mass index (kg/m^2)
Surgery Information	Duration of operation (mins)
PACU Pain Score	Patient's last PACU pain score recorded (NRS 0 to 10)
	Standard deviation of all PACU pain scores recorded
Postoperative pain score (at rest)	Maximum postoperative pain score at the 6th hour (NRS 0 to 10)
	Maximum postoperative pain score at the 9th hour (NRS 0 to 10)
	Mean of each hour's maximum postoperative pain score at the 0th to 12th hour
	Standard deviation of each hour's maximum postoperative pain score at the 0th to 12th hour
	Whether more than one pain score was recorded within an hour at any point of time (yes/no)
	Mean time difference between each postoperative pain score recorded (mins)
	Minimum time difference between each postoperative pain score recorded (mins)
	Time taken from last recorded postoperative pain score to the 12th hour prediction point (mins)
	Time taken from the 2nd last to the last postoperative pain score recorded (mins)
	Time taken from the 3rd last to 2nd last postoperative pain score recorded (mins)
Postoperative pain score (on movement)	Last recorded postoperative pain score before prediction (NRS 0 to 10)
	Minimum postoperative pain score (NRS 0 to 10)
	Maximum postoperative pain score at the 1st hour (NRS 0 to 10)
	Time taken from the 1st to 2nd postoperative pain score recorded (mins)
	Time taken from the 2nd to 3rd postoperative pain score recorded (mins)
	Time taken from the 3rd to 4th postoperative pain score recorded (mins)
	Time taken from the 4th last to 3rd last postoperative pain score recorded (mins)

NRS, numerical rating scale; PACU, post-anaesthesia care unit; SULOV, Searching for Uncorrelated List of Variables.

Table 2
The hyperparameters of the six machine learning algorithms used in this study.

Machine Learning Models	Hyperparameters	Definition	Defined Parameters
Ridge regression	Solver	The algorithm used to optimise the loss function	newton-cg, lbfgs, liblinear, sag, saga
	C-value	Constant	0.001–100
LASSO	Solver	The algorithm used to optimise the loss function	newton-cg, lbfgs, liblinear, sag, saga
	C-value	Constant	0.001–100
Elastic net	L1 ratio value	The ratio between L1 (LASSO) and L2 (Ridge) regularization	0.1–0.9
	Solver	The algorithm used to optimise the loss function	newton-cg, lbfgs, liblinear, sag, saga
	C-value	Constant	0.001–100
Random Forest	n_estimators	The number of decision trees	62–700
	max_features	The number of features to consider when looking for the best split	sqrt, log2, all features
	max_depth	The maximum depth of each tree	3–7
	min_samples_leaf	The minimum number of samples required to form a leaf node	2–11
	min_samples_split	The minimum number of samples required to split an internal node	2–11
	Bootstrap	Whether bootstrap samples are used when building trees	True, False
XGBoost	reg_lambda	The L2 regularization term on weight	0–30
	reg_alpha	The L1 regularization term on weights	0–30
	gamma	The minimum loss reduction required to make a further partition on a leaf node	0.1–0.4
	max_depth	The maximum depth of each tree	1–8
	n_estimators	The number of boosting rounds or trees to build	50–1500
	learning_rate	The step size shrinkage used in update to prevent overfitting	0.01–0.2
LightGBM	num_leaves	The maximum number of leaves in each tree	20–50
	max_depth	The maximum depth of each tree	1–6
	num_iterations	The number of iterations of boosting rounds	25–100
	learning_rate	The step size shrinkage used in update to prevent overfitting	0.03–0.2
	Boosting	The type of boosting to use	gbdt, dart
	data_sample_strategy	The strategy for sampling data	bagging, goss

LASSO, Least Absolute Shrinkage and Selection Operator.

and negative predictive value (NPV) were also calculated using optimal cut-off values determined by the intersection between sensitivity and specificity.

3. Results

A total of 7251 caesarean patients were administered postoperative spinal morphine within 48 h from August 2019 to August 2022; 690 were excluded due to missing primary outcome (Fig. 2). The remaining 6561 patients were included in the study dataset, with 521 (7.9 %) experiencing a maximum postoperative pain score of 3 or more on movement at the 13th to 24th hour after spinal morphine administration (Table 3).

Eighty percent of the study cohort (n = 5248) were randomly selected for our training cohort, of which 7.8 % (n = 403) experienced maximum postoperative NRS pain score of 3 or more on movement at the 13th to 24th hour after spinal morphine administration. The predictive performance of the models was validated using the remaining 20 % of the study cohort (n = 1313), of whom 8.9 % (n = 118) experienced a maximum postoperative NRS pain score of 3 or more on movement at the 13th to 24th hour after spinal morphine administration.

A comparison of the performance of the six different risk stratification predictive models is presented in Table 4. All models except LightGBM (AUC: 0.691) scored an AUC above 0.7, with ridge regression achieving the highest AUC of 0.719. To demonstrate the impact of feature selection optimization using the SULOV-recursive method, we compared the model performance on the selected 23 features (Table 1; Appendix S1; Fig. 3A) against the baseline models without feature selection (Fig. 3B, which is based on the full 120 features; Appendix S2). Following the feature selection optimization using the SULOV-recursive method, the AUCs of all the models

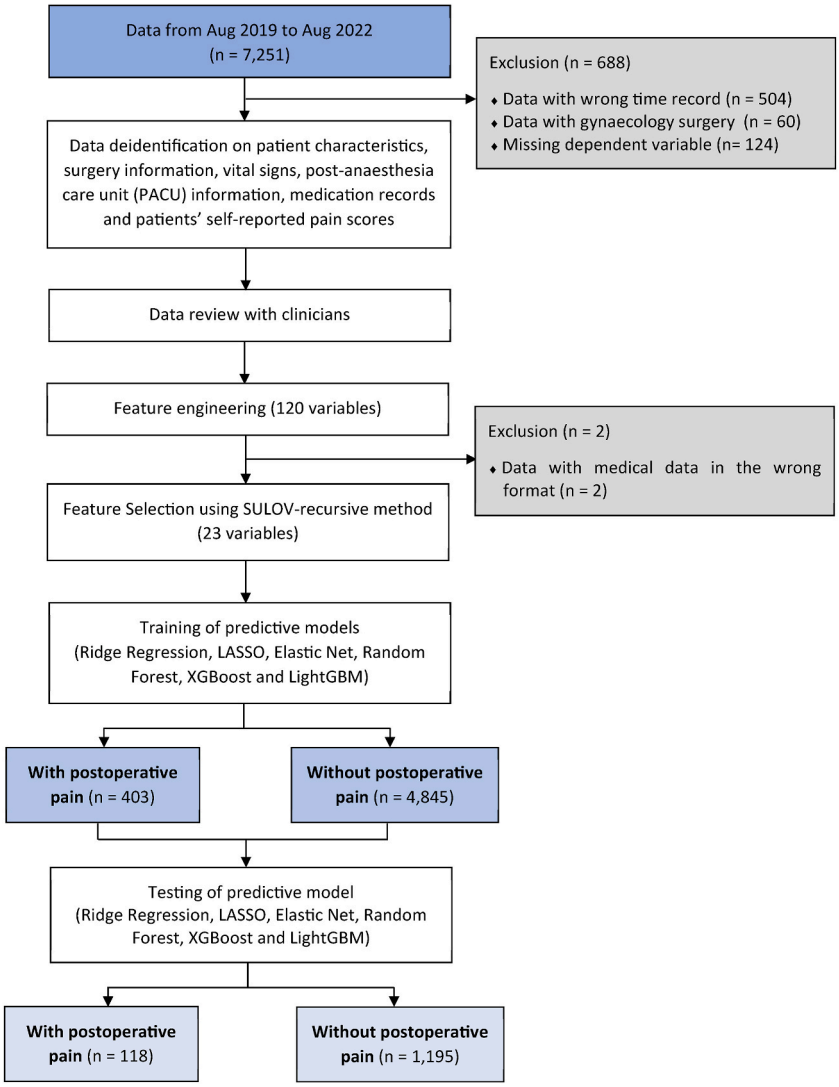


Fig. 2. Study workflow.

Table 3
Demographic and clinical characteristics. Values are mean (SD), median (IQR [range]) or number (proportion).

Variable	Training cohort (n = 5248)		Validation cohort (n = 1313)	
	Postoperative pain		Postoperative pain	
	Yes (n = 403)	No (n = 4845)	Yes (n = 118)	No (n = 1195)
Maternal age (y), median (IQR)	30.0 (7.0 [30.0, 37.0])	33.0 (6.0 [30.0, 36.0])	33.6 (6.0 [30.0, 36.0])	33.0 (7.0 [30.0, 37.0])
Race				
Chinese	132 (32.8 %)	2248 (46.4 %)	36 (30.5 %)	548 (45.9 %)
Malay	120 (29.8 %)	1040 (21.5 %)	30 (29.7 %)	268 (22.4 %)
Indian	78 (19.4 %)	614 (12.7 %)	19 (16.1 %)	147 (12.3 %)
Others	73 (18.1 %)	945 (19.5 %)	28 (23.7 %)	232 (19.4 %)
Height (cm)	157.6 (5.5)	157.5 (6.8)	157.6 (5.4)	157.7 (6.3)
Weight (kg)	76.3 (15.4)	74.5 (13.7)	76.4 (13.2)	75.3 (13.5)
Body mass index (kg/m^2), mean (SD)	30.7 (5.9)	30.0 (6.1)	30.7 (5.1)	30.3 (5.3)
Preoperative analgesia[a]				
Paracetamol	1 (0.3 %)	11 (0.2 %)	0 (0)	1 (0.1 %)
Mefenamic acid	4 (1.0 %)	27 (0.6 %)	0 (0)	11 (0.9 %)
Tramadol	1 (0.3 %)	0.0 (0)	1 (0.9 %)	1 (0.1 %)
Average dosage if preoperative analgesia was administered				
Paracetamol (mg)	975.0 (0.0)	1985.0 (0.0)	0.0 (0.0)	1000.0 (0.0)
Mefenamic acid (mg)	500.0 (0.0)	833.3 (561.1)	0.0 (0.0)	590.9 (287.5)
Tramadol (mg)	100.0 (0.0)	0.0 (0.0)	50.0 (0.0)	50.0 (0.0)
Type of admission				
Emergency delivery	176 (43.7 %)	1864 (38.5 %)	52 (44.1 %)	462 (38.7 %)
Elective delivery	227 (56.3 %)	2981 (61.3 %)	66 (55.9 %)	733 (61.3 %)
Duration of operation (mins), median (IQR)	44.0 (26.0 [33.0, 59.0])	43.0 (22.0 [33.0, 55.0])	43.5 (20.0 [35.0, 55.0])	43.0 (21.0 [34.0, 55.0])
Procedural code[b]				
Uncomplicated caesarean	349 (86.6 %)	4217 (87.0 %)	98 (83.1 %)	1014 (84.9 %)
Uncomplicated caesarean and tubal ligation	38 (9.4 %)	453 (9.4 %)	14 (11.9 %)	124 (10.4 %)
Complicated caesarean	10 (2.5 %)	131 (2.7 %)	5 (4.2 %)	48 (4.0 %)
Complicated caesarean and tubal ligation	3 (0.7 %)	33 (0.7 %)	1 (0.9 %)	8 (0.7 %)
Caesarean with hysterectomy	1 (0.3 %)	7 (0.1 %)	0 (0)	0 (0)
Others	2 (<0.1 %)	4 (0.1 %)	0 (0)	1 (0.1 %)
Surgery time				
0830 to 2029	331 (83.1 %)	3947 (81.5 %)	96 (81.4 %)	960 (80.3 %)
2030 to 0829	72 (17.9 %)	898 (18.5 %)	22 (18.6 %)	235 (19.7 %)
Average PACU pain score (NRS 0–10), median (IQR)	1.0 (0.1 [1.0, 1.1])	1.0 (0.0 [1.0, 1.0])	1.0 (0.0 [1.0, 1.0])	1.0 (0.0 [1.0, 1.0])
Maximum PACU pain score (NRS 0–10), median (IQR)	1.0 (0.5 [1.0, 1.5])	1.0 (0.0 [1.0, 1.0])	1.0 (0.0 [1.0, 1.0])	1.0 (0.0 [1.0, 1.0])
Average postoperative pain score at the 0th to 12th hour (NRS 0–10, at rest), median (IQR)	0.5 (0.8 [0.1, 0.9])	0.1 (0.5 [0.0, 0.5])	0.6 (0.8 [0.1, 0.9])	0.1 (0.5 [0.0, 0.5])
Maximum postoperative pain score at the 0th to 12th hour (NRS 0–10, at rest), median (IQR)	1.0 (1.0 [1.0, 2.0])	1.0 (1.0 [0.0, 1.0])	1.0 (1.0 [1.0, 2.0])	1.0 (1.0 [0.0, 1.0])
Average postoperative pain score at the 0th to 12th (NRS 0–10, on movement), median (IQR)	1.0 (1.1 [0.5, 1.6])	0.6 (0.8 [0.2, 1.0])	1.1 (1.0 [0.6, 1.6])	0.6 (0.9 [0.2, 1.1])
Maximum postoperative pain score at the 0th to 12th hour (NRS 0–10, on movement), median (IQR)	2.0 (2.0 [1.0, 3.0])	1.0 (1.0 [1.0, 2.0])	1.0 (1.0 [1.0, 2.0])	1.0 (1.0 [1.0, 2.0])
Nausea	10 (2.5 %)	163 (3.4 %)	3 (2.5 %)	45 (3.8 %)
Pruritus	24 (6.0 %)	148 (3.1 %)	0 (0)	38 (3.2 %)
Vomit	12 (3.0 %)	170 (3.5 %)	4 (3.4 %)	39 (3.3 %)
SpO_2 < 95 %	9 (2.2 %)	49 (1.0 %)	3 (2.5 %)	15 (1.3 %)
Respiration rate <10 breaths per min at any point of time	0 (0)	16 (0.3 %)	0 (0)	1 (0.1 %)
Supplementary oxygen therapy (nasal prong/nasal cannula) at any point of time	37 (9.2 %)	169 (3.5 %)	8 (6.8 %)	49 (4.1 %)
Average postoperative pain score at the 13th to 24th hour (NRS 0–10, at rest), median (IQR)	0.4 (0.5 [0.2, 0.7])	0.0 (0.1 [0.0, 0.1])	0.5 (0.5 [0.2, 0.7])	0.0 (0.1 [0.0, 0.1])
Maximum postoperative pain score at the 13th to 24th hour (NRS 0–10, at rest), median (IQR)	2.0 (2.0 [1.0, 3.0])	0.0 (1.0 [0.0, 1.0])	2.0 (2.0 [1.0, 3.0])	0.0 (1.0 [0.0, 1.0])
Average postoperative pain score at the 13th to 24th hour (NRS 0–10, on movement), median (IQR)	1.0 (0.7 [0.6, 1.3])	0.1 (0.4 [0.0, 0.4])	0.9 (0.7 [0.6, 1.3])	0.1 (0.4 [0.0, 0.4])
Maximum postoperative pain score at the 13th to 24th hour (NRS 0–10, on movement), median (IQR)[c]	3.0 (0.0 [3.0, 3.0])	1.0 (1.0 [0.0, 1.0])	3.0 (0.0 [3.0, 3.0])	1.0 (1.0 [0.0, 1.0])
Average body temperature (°C), median (IQR)	36.9 (0.4 [36.8, 37.1])	36.9 (0.3 [36.8, 37.1])	37.0 (0.3 [36.8, 37.1])	36.9 (0.3 [36.8, 37.1])
Maximum body temperature (°C), median (IQR)	37.2 (0.5 [36.9, 37.4])	37.1 (0.5 [36.9, 37.4])	37.2 (0.4 [37.0, 37.4])	37.1 (0.3 [37.0, 37.3])
Postoperative analgesia				

(*continued on next page*)

Table 3 (*continued*)

Variable	Training cohort (n = 5248)		Validation cohort (n = 1313)	
	Postoperative pain		Postoperative pain	
	Yes (n = 403)	No (n = 4845)	Yes (n = 118)	No (n = 1195)
Paracetamol	9 (2.2 %)	46 (1.0 %)	3 (2.5 %)	13 (1.1 %)
Mefenamic acid	322 (79.9 %)	4440 (91.6 %)	101 (85.6 %)	1131 (94.6 %)
Tramadol	0 (0)	2 (<0.1 %)	0 (0)	0 (0)
Average dosage if postoperative analgesia was administered				
Paracetamol (mg)	1000.0 (0.0)	997.39 (9.9)	1000.0 (0.0)	1000.0 (0.0)
Mefenamic acid (mg)	508.5 (69.1)	513.85 (86.1)	517.33 (101.1)	511.27 (75.3)
Tramadol (mg)	0.0 (0.0)	50.0 (0.0)	0.0 (0.0)	0.0 (0.0)

IQR, interquartile range; NRS, numerical rating scale; PACU, post-anaesthesia care unit; SD, standard deviation; SpO_2, oxygen saturation.

[a] The medication data was taken within the first 12 h only.

[b] Complicated caesarean refers to case of abnormally invasive placenta, massive postpartum haemorrhage >1.5 L, preterm caesarean <34 weeks, or abruptio placenta [30].

[c] The primary outcome of the study.

Table 4
Comparison of performance of different models.

Feature Selection	Model	AUC (99 % CI)	Cut-off	Sensitivity (99 % CI)	Specificity (99 % CI)	PPV (99 % CI)	NPV (99 % CI)
With feature selection (Top 23 features)	Ridge regression	0.719 (0.659–0.775)	0.51	0.653 (0.550–0.748)	0.705 (0.674–0.735)	0.179 (0.136–0.224)	0.954 (0.936–0.968)
	LASSO	0.717 (0.654–0.778)	0.52	0.644 (0.538–0.740)	0.685 (0.655–0.715)	0.168 (0.128–0.210)	0.951 (0.933–0.967)
	Elastic net	0.718 (0.653–0.777)	0.53	0.695 (0.597–0.796)	0.644 (0.613–0.677)	0.162 (0.124–0.201)	0.955 (0.937–0.972)
	Random Forest	0.704 (0.639–0.759)	0.53	0.636 (0.529–0.735)	0.657 (0.627–0.687)	0.155 (0.116–0.195)	0.948 (0.930–0.965)
	XGBoost	0.715 (0.652–0.771)	0.52	0.644 (0.537–0.742)	0.653 (0.620–0.684)	0.155 (0.117–0.194)	0.949 (0.930–0.966)
	LightGBM	0.691 (0.624–0.752)	0.54	0.644 (0.540–0.743)	0.662 (0.633–0.693)	0.158 (0.120–0.198)	0.950 (0.931–0.966)
Without feature selection (Full 120 features)	Ridge regression	0.649 (0.595–0.699)	0.52	0.653 (0.554–0.745)	0.645 (0.615–0.677)	0.154 (0.116–0.192)	0.950 (0.932–0.966)
	LASSO	0.669 (0.613–0.722)	0.53	0.661 (0.553–0.758)	0.676 (0.646–0.706)	0.168 (0.128–0.210)	0.953 (0.936–0.969)
	Elastic net	0.649 (0.596–0.703)	0.51	0.661 (0.559–0.760)	0.636 (0.606–0.668)	0.152 (0.115–0.191)	0.950 (0.931–0.967)
	Random Forest	0.653 (0.599–0.702)	0.54	0.653 (0.550–0.746)	0.654 (0.622–0.685)	0.157 (0.117–0.197)	0.950 (0.932–0.967)
	XGBoost	0.620 (0.567–0.674)	0.52	0.627 (0.526–0.726)	0.613 (0.579–0.644)	0.138 (0.103–0.174)	0.943 (0.925–0.962)
	LightGBM	0.637 (0.585–0.688)	0.52	0.627 (0.523–0.726)	0.646 (0.613–0.678)	0.149 (0.112–0.188)	0.946 (0.928–0.963)

AUC, area under the curve; LASSO, Least Absolute Shrinkage and Selection Operator; NPV, Negative predictive value; PPV, Positive predictive value.

increased by 7 %–13 % (Table 4).

To further analyse the findings of the top four performing models (ridge regression, LASSO, Elastic net, and XGBoost), we assessed the impact of individual features on each model by referencing the coefficients of the regularized regression models and the variable importance scores of the ensemble learning algorithms (Appendix S3). Notably, these four modelling techniques identified the same top three variables that contributed to predictive performance: the last recorded postoperative pain score (on movement) before the prediction point, the standard deviation of each hour's maximum postoperative pain score (at rest) at the 0^{th} to 12th hour, and the mean of each hour's maximum postoperative pain score (at rest) at the 0^{th} to 12th hour. Ridge regression utilizes the maximum number of variables by incorporating all 23 selected features, whereas LASSO regression employs the fewest, specifically retaining only the same top three variables as the other three modelling techniques.

4. Discussion

We reported significant pain in 8 % of the study cohort, with a maximum postoperative pain score of 3 or more on movement at the 13th to 24th hour after spinal morphine administration during caesarean delivery. The predictive performances of six different risk stratification models were compared using 120 clinical variables extracted or generated from the medical records: Ridge regression, LASSO, Elastic net, Random Forest, XGBoost, and LightGBM. Four of the six modelling techniques (ridge regression, LASSO, Elastic net,

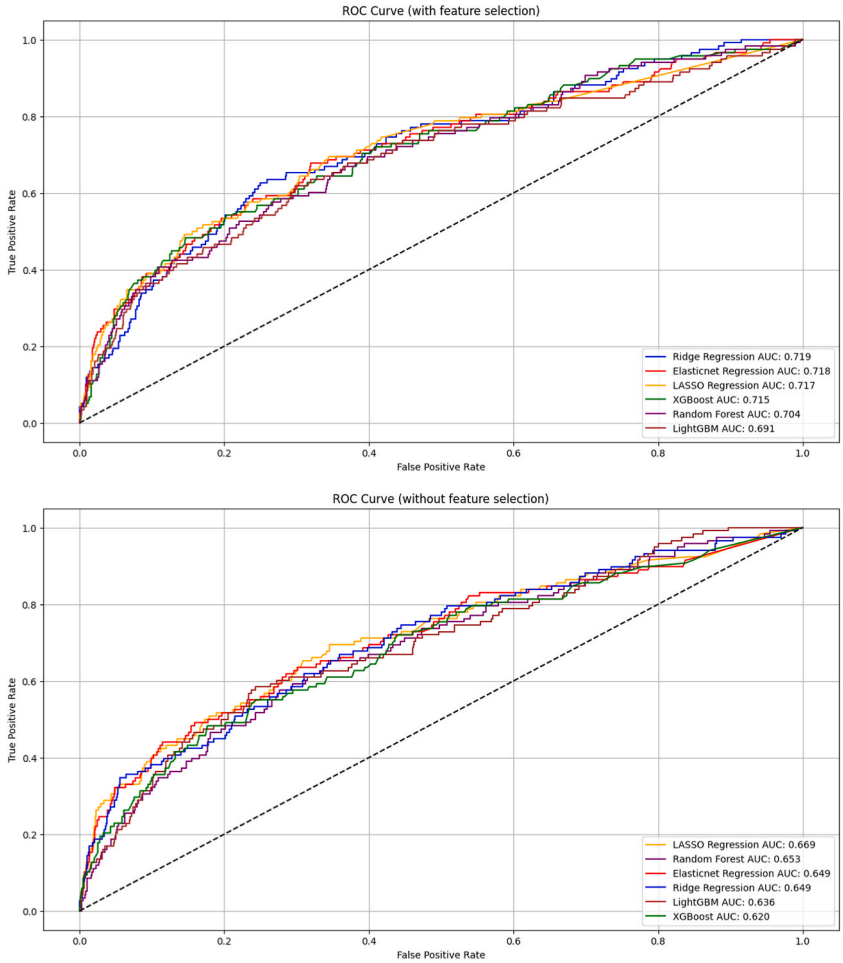


Fig. 3. Comparison of ROC curve across models (A) with; and (B) without feature selection.

and XGBoost) demonstrated similar performance in terms of AUC, sensitivity, specificity, PPV and NPV when predicting significant postoperative pain after caesarean delivery. Reducing the number of predictor variables to 23 using the SULOV-recursive method enhanced the accuracy of the tested algorithms. Ridge regression exhibited the best performance with both the full and selected feature sets.

There is increasing interest in applying machine learning techniques in acute postoperative pain settings [5,6,19]. With a sample size of 8071 patients undergoing non-obstetric, non-ambulatory surgeries, Tighe et al. extracted 796 clinical variables from a surgical database for further model comparison, including demographic data, comorbidities, outpatient analgesia, and surgical information [6]. By predicting the moderate-to-severe acute postoperative pain at the 0^{th} to 24th hour, the authors showed that LASSO algorithm had the highest area under the receiver-operating curve (ROC) of 0.704 when all 796 variables were used, while logistic regression performed poorly with an ROC of 0.500 [6]. Interestingly, the study did not consider the use of perioperative analgesia despite extracting more variables than our study. In addition, the study lacked data on the mean and SD of pain scores, as well as the time differences between the pain scores. One plausible explanation for this could be the occurrence of over-fitting issues, stemming from an excessive number of variables and hence leading to increased multicollinearity [20]. This further emphasizes the importance of feature selection in reducing model overfitting and enhancing the predictive power of the algorithms. In our case, we observed an AUC increase of 7 %–13 % across all models after optimizing the feature selection [10].

Similarly, Davoudi et al. also investigated the model predictive performance for acute postoperative pain from 14,263 patients who underwent orthopaedic surgery. The authors utilized a similar gradient boosting algorithms (XGBoost and LightGBM, CatBoost etc.) with a reported AUC of 0.71, albeit with potential biases related to age, race, area deprivation index, and type of insurance [21]. Notably, both our study and Davoudi et al. did not collect any preoperative psychological data, as this information is not universally collected and often takes the form of questionnaires. In a previous prospective cohort study ($n = 217$), we found that only anxiety about upcoming surgery was significantly associated with moderate-to-severe acute postoperative pain at the 24th hour [22]. In a similar vein, Kalkman et al. incorporated psychological data such as quality of life and anxiety to enhance model performance [23]. Considering that pain is a multifactorial experience, incorporating psychological information could improve our understanding of

patient experiences and enhance model performance in pain risk stratification.

We found that ridge regression outperformed other conventional modelling techniques in this study, especially when only selected features were used in the model. This may be attributed to its relation to ensemble learning algorithms with simpler structured data, and the linear relationships between variables and the outcome of postoperative pain after caesarean delivery. Our ridge regression model included the top 23 features selected for model development, encompassing variables such as age, height, BMI, duration of surgery, pain scores, time, and intervals at which pain scores were documented. In contrast, the other regularized logistic regression model such as LASSO included only three variables that contributed to the predictive model's performance. This is likely due to its built-in ability to perform feature selection by penalizing the absolute value of each individual coefficient and shrinking the coefficients of irrelevant variables to 0 [12].

Among the features selected via the SULOV-recursive method, the last recorded postoperative pain score, the mean and SD of each hour's maximum postoperative pain score (at rest) were the top three features, particularly the postoperative pain scores reported by the patients prior to the 13th hour after spinal morphine administration. Notably, these three variables also overlapped with the top three predictors for ridge regression, LASSO, Elastic net, and XGBoost, which underscores the importance assessing and monitoring inpatient pain scores in high-risk patients. However, this also highlights the underlying issue in patients with cognitive or verbal difficulties in communicating their pain scores, urging the need to progress towards better automated pain assessments.

We acknowledge that the significant postoperative pain defined in this study was measured using NRS over a finite period; yet the subjective and dynamic nature of pain, including our primary outcome, may not be applicable to every patient. Our findings also indicated the time differences between the pain scores, both with and without considering the time elapsed, as part of the selected features during model development. Rahman et al. previously coined the term "pain volatility" to describe the average absolute difference between two consecutive pain scores at the respective observation points without accounting for the time elapsed [24]. In contrast to many studies utilizing retrospective medical records, some research employed mobile apps to collect patients' demographic and pain information for further model development. The results showed that Random Forest achieved the best performance of 70 % accuracy in high volatility, which reflects great uncertainty in pain experiences. Nonetheless, the study was based on self-reported outcome of persistent pain at six months in community users, and the findings may not be applicable to inpatient settings for better clinical decision making.

4.1. Strength and limitations

The strength of the study lies in the large sample size gathered over three years, which enhances the reliability and generalizability of the models in contemporary clinical settings. The variables selected for modelling are clinically relevant and commonly collected across different healthcare institutions, facilitating easier future integration into clinical practice. Additionally, the use of six different predictive modelling techniques provides a comprehensive evaluation of the accuracy and relevance of the developed models with the selected features.

We acknowledge several limitations in this study. First, we utilized retrospective patient data from a single study site, which limited our ability to perform personalized treatment at the individual level. The selected top 23 features comprised only three pre-delivery factors (age, height, BMI) to help in planning pain management. Important maternal (e.g., gestational weeks, gravida, parity, history of past caesarean delivery), intraoperative (e.g., neuraxial technique, analgesia) characteristics, psychological characteristics (previous experience, expectation, anxiety [24]), and delivery outcomes (e.g., Apgar scores, blood loss, complications) were not captured systematically in our enterprise platform and hence could not be extracted for this study. There was also a lack of vocal information and natural language processing (NLP) from both videos and case notes during model development, which has been reported to improve model performance on acute postoperative pain in depressed patients [25]. However, despite these challenges, we were still able to generate meaningful data for risk prediction. Future work will focus on addressing the data domains that were lacking in this study by enhancing the electronic medical record system to further improve the predictive models.

Second, we reported an incidence of significant postoperative pain of 8 %, which is comparably lower than other reports [26]. It is important to note that our study was done primarily in Asian population, which may exhibit different pain thresholds and perceptions compared to other studies [27]. In addition, the local clinical practice discourages the prescription of opioid as the first-line treatment for acute and chronic pain [28]; instead, the use of alternative pharmacological and non-pharmacological modalities is firstly maximized for pain relief. This results in a significant proportion of the population being opioid-naïve with varying pain responses. Furthermore, our institution adopts the effective use of intrathecal morphine as a postoperative analgesia for all patients who underwent Caesarean surgery, which could also contribute to the lower incidence of significant postoperative pain observed in our study. This low incidence also implies an imbalance in the pain datasets as compared to those without pain. Thus, although the AUCs in the presented models are high, their precisions are considered relatively low as the analysis focused on the sensitivity and specificity of the developed models using imbalanced datasets. Lastly, the best-performing model on ridge regression represents a linear relationship between the features and the significant postoperative pain. There may be other forms of non-linear relationships not measured in this study, which could be important for evaluating the pain profile.

4.2. Clinical significance and future work

Notably, the models presented in this study included many features related to postoperative pain scores. At our institution, clinical information including vital signs and pain scores, is entered and displayed across different sections of the platform. This fragmentation renders pain prediction and personalization difficult as the availability of pain scores alone does not allow healthcare professionals to

access information pertinent to personalized pain experience. Furthermore, the acute pain ward rounds are routinely performed on postoperative day one (24 h after surgery), while the pain score assessments are performed hourly during the first 24 h. Therefore, a risk stratification dashboard that utilizes the information prior to the pain rounds will help to identify elective surgery patients at risk of significant pain in the 13th – 24th hour - time points that typically fall after office hours with limited manpower. The current morphine dosage used at our institution is considered low-dose analgesia with a range of approximately 10–27 h until the first request for additional analgesia [29]. With the proposed risk models, we will be able to stratify the patients immediately after the surgery to enable earlier review and interventions during pain rounds. However, some patients (8 %) with significant pain may still require additional analgesics even after identification using this risk stratification strategy. In these cases, first-line treatments such as paracetamol and mefenamic acid are usually administered, while those escalating to severe or prolonged pain will receive tramadol as a second-line treatment.

Thus, the ability to objectively and reliably identify patients at increased risk of significant postoperative pain in a timely manner will address critical gaps in post-caesarean pain management. This will allow clinicians to prioritize patient assessment and management at elevated risk of significant pain, and potentially reduce the risk of pain-related morbidities to improve patient outcomes and reduce healthcare costs. By deploying the risk stratification model in clinical practice, we could reduce the time needed to identify patients with significant pain and minimize disruptions to clinical workflow caused by ad hoc pain reviews. In addition, the developed risk stratification model can be scaled up and integrated with our hospital's electronic medical record system, for instance a dashboard that efficiently and intuitively conveys high-risk information to healthcare professionals, helping them prioritize high-risk patients and streamline their clinical workflows. Other work will also focus on addressing the data domains that were lacking in this study by enhancing the electronic medical record system to further improve the predictive models.

5. Conclusion

Employing the ridge regression model is the most effective way to predict significant postoperative pain after caesarean delivery. Future plans include prospective validation and integration into healthcare systems to benefit both caesarean patients and healthcare professionals by streamlining the clinical workflow with risk stratification, allowing for better responses to patients at high risk of significant postoperative pain, which ultimately enhances patient outcomes and experience.

CRediT authorship contribution statement

Chin Wen Tan: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Juan Zhen Koh:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Hanwei Jin:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Nian-Lin Reena Han:** Writing – review & editing, Software, Investigation, Data curation, Conceptualization. **Shang-Ming Cheng:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Andy Wee An Ta:** Writing – review & editing, Visualization, Supervision, Software, Investigation, Conceptualization. **Han Leong Goh:** Writing – review & editing, Visualization, Supervision, Software, Investigation, Conceptualization. **Ban Leong Sng:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Presentation

None.

Trial registration

Not applicable.

Data and code availability

Data will be made available on request.

Financial support and sponsorship

This study was supported by the SingHealth Duke-NUS Anesthesiology and Perioperative Sciences Academic Clinical Program Philanthropy Grant 2022. The funder had no role in the study design, data collection and analysis, interpretation of data, or the manuscript writing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Chin Wen Tan is a section editor of Heliyon, Women's Health section. All other authors declare no competing interests.

Acknowledgments relating to this article

Assistance with the study: The authors would like to thank Dr Paul Tan Hon Sen (Consultant) and Ms Agnes Teo (Senior Clinical Research Coordinator) for their administrative support in this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e40602.

References

[1] B.H. Gamez, A.S. Habib, Predicting severity of acute pain after cesarean delivery: a narrative review, Anesth. Analg. 126 (2018) 1606–1614.
[2] S.A. Schug, G.M. Palmer, D.A. Scott, M. Alcock, R. Halliwell, J.F. Mott, Acute Pain Management: Scientific Evidence, fifth ed., Acute pain management scientific evidence working group of the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine, Melbourne, 2020.
[3] J.C. Eisenach, P.H. Pan, R. Smiley, P. Lavand'homme, R. Landau, T.T. Houle, Severity of acute pain after childbirth, but not type of delivery, predicts persistent pain and postpartum depression, Pain 140 (2008) 87–94.
[4] World Health Organization, Caesarean Section Rates Continue to Rise, amid Growing Inequalities in Access, World Health Organization, Geneva, June 16, 2021. https://www.who.int/news/item/16-06-2021-caesarean-section-rates-continue-to-rise-amid-growing-inequalities-in-access. (Accessed 26 June 2024).
[5] A.A. Nair, M.A. Velagapudi, J.A. Lang, et al., Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients, PLoS One 15 (2020) e0236833.
[6] P.J. Tighe, C.A. Harle, R.W. Hurley, H. Aytug, A.P. Boezaart, R.B. Fillingim, Teaching a machine to feel postoperative pain: combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain, Pain Med. 16 (2015) 1386–1401.
[7] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, Ann. Intern. Med. 162 (2015) 55–63.
[8] L.L. Xu, C. Wang, C.M. Deng, et al., Efficacy and safety of esketamine for supplemental analgesia during elective cesarean delivery: a randomized clinical trial, JAMA Netw. Open 6 (2023) e239321.
[9] M. Khushi, K. Shaukat, T.M. Alam, et al., A comparative performance analysis of data resampling methods on imbalance medical data, IEEE Access 9 (2021) 109960–109975.
[10] J. Brownlee, Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python, first ed., Machine Learning Mastery, 2020.
[11] C. Shu, C. Zheng, D. Luo, J. Song, Z. Jiang, L. Ge, Acute ischemic stroke prediction and predictive factors analysis using hematological indicators in elderly hypertensives post-transient ischemic attack, Sci. Rep. 14 (2024) 695.
[12] R. Tibshirani, Regression shrinkage and selection via the LASSO, J R Stat Soc Series B Stat Methodol 58 (1996) 267–288.
[13] R. Kohavi, D.H. Wolpert, Bias plus variance decomposition for zero-one loss functions, in: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, 1996, pp. 275–283.
[14] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1970) 55–67.
[15] H. Zou, T. Hastie, Regularization and variable selection via the Elastic net, J R Stat Soc Series B Stat Methodol 67 (2005) 301–320.
[16] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[17] G. Ke, Q. Meng, T. Finley, et al., LightGBM: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. (2017) 3149–3157.
[18] J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2009) 569–575.
[19] L. Juwara, N. Arora, M. Gornitsky, P. Saha-Chaudhuri, A.M. Velly, Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning, Int. J. Med. Inf. 141 (2020) 104170.
[20] S. Ray, A quick review of machine learning algorithms, in: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, 2019, pp. 35–39.
[21] A. Davoudi, R. Sajdeya, R. Ison, et al., Fairness in the prediction of acute postoperative pain using machine learning models, Front Digit Health 4 (2023) 970281.
[22] J.J. Chan, C.W. Tan, C.T. Yeam, et al., Risk factors associated with development of acute and sub-acute post-cesarean pain: a prospective cohort study, J. Pain Res. (2020) 2317–2328.
[23] C.J. Kalkman, K. Visser, J. Moen, G.J. Bonsel, D.E. Grobbee, K.G. Moons, Preoperative prediction of severe postoperative pain, Pain 105 (2003) 415–423.
[24] Q.A. Rahman, T. Janmohamed, M. Pirbaglou, et al., Defining and predicting pain volatility in users of the manage my pain app: analysis using data mining and machine learning methods, J. Med. Internet Res. 20 (2018) e12001.
[25] A. Parthipan, I. Banerjee, K. Humphreys, et al., Predicting inadequate postoperative pain management in depressed patients: a machine learning approach, PLoS One 14 (2019) e0210575.
[26] B.C. Demilew, N. Zurbachew, N. Getachew, G. Mekete, D.T. Lema, Prevalence and associated factors of postoperative acute pain for mothers who gave birth with cesarean section: a systematic review and meta-analysis, Pain Manag. Nurs. (2024) in press.
[27] H.J. Kim, G.S. Yang, J.D. Greenspan, et al., Racial and ethnic differences in experimental pain sensitivity: systematic review and meta-analysis, Pain 158 (2017) 194–211.
[28] Singapore Ministry of Health, National guidelines for the safe prescribing of opioids (Apr 2021) librariesprovider5/default-document-library/national-guidelines-for-the-safe-prescribing-of-opioids-2021.pdf.
[29] P. Sultan, S.H. Halpern, E. Pushpanathan, S. Patel, B. Carvalho, The effect of intrathecal morphine dose on outcomes after elective cesarean delivery: a meta-analysis, Anesth. Analg. 123 (2016) 154–164.
[30] Singapore Ministry of Health, List of Benchmarks (June 14, 2023). https://www.moh.gov.sg/docs/librariesprovider5/default-document-library/full-list-of-fee-benchmarks_230615-(2).pdf. (Accessed 18 June 2024).