

Application of Artificial Intelligence Models in Predicting Clinical Outcomes after Nerve Block in Orthopedic Surgery

Jorge Rivadulla Brey, Jerónimo Pardo-Rodríguez, Laura García-Reza, Vanessa Aguiar-Pulido, and Miguel Angel Pereira Loureiro

Universidade da Coruña, CITIC, TALIONIS Research Group, 15071 A Coruña, Spain
University of A Coruña, University of Miami, Weill Cornell Medicine
Hospital Álvaro Cunqueiro

Correspondence: jorge.rivadulla@udc.es

DOI: <https://doi.org/10.17979/spu.23.c38>

Abstract: This project aims to analyze the effect of the anesthesiology technique of nerve block on the postoperative outcomes of patients undergoing hip and knee surgery. Using the collected data, which includes variables such as pain levels, medication intake, and satisfaction with the procedure, the goal is to develop an artificial intelligence model capable of determining which approach, using a nerve block or not, offers better clinical results. The main objective of this project is to demonstrate that the use of nerve blocks contributes to a less painful recovery and a lower consumption of analgesics.

1 Introduction

Regional anesthesia, especially through nerve blocks for hip and spine interventions, has been shown to be an effective strategy for improving postoperative pain control, reducing the need for opioids, and promoting early mobilization. However, the efficacy of these anesthetic blocks can vary significantly between patients depending on multiple clinical, demographic, or surgical factors. A thorough understanding of these factors can help optimize both patient selection and the planning of individualized anesthesia procedures, thereby improving clinical outcomes.

In recent years, supervised machine learning techniques have become promising tools for predicting clinical outcomes and extracting relevant information from medical datasets. Examples include the work of Li et al. (2022) applied to preoperative risks in orthopedics, and Zhou et al. (2024), where explainable models were developed to anticipate transfusions in hip fracture surgery, identifying clinically relevant variables.

Despite these advances, challenges remain when data were not originally designed to predict a variable of interest. Small sample sizes, heterogeneity, and missing values can promote overfitting Silvey and Liu (2024). Another important point is that the estimation of variable importance can be biased by correlations or measurement scales, as documented in studies such as Wies et al. (2023) and Strobl et al. (2007).

Furthermore, designing algorithms with reproducible data preprocessing is essential to ensure robust models, as highlighted in recent research conducted in clinical settings with small datasets, such as Wang and Pai (2023), who advocate the use of artificial datasets to increase model robustness.

Within this framework, the present work aims to apply supervised classification models, such as Random Forest, Lasso regression, and Support Vector Machines, to real clinical data

in order to evaluate their performance, identify relevant variables, and discuss their strengths and limitations. Based on these considerations, the following objectives were defined for this study.

2 Objectives

The present work has as its objective to explore the application of supervised machine learning techniques to a real clinical dataset with the purpose of identifying factors associated with the success of the block in hip and spine surgeries. In general terms, the main objective of this work is to evaluate the viability of employing classification models and methods of variable importance analysis to predict the occurrence of an anesthetic block in situations in which the data have not been generated with that objective. This approach seeks, in addition to assessing the predictive performance of the different algorithms, to examine their capacity to provide useful interpretations about which characteristics of the patients and of the anesthetic procedure have the greatest weight in the patient enjoying a better postoperative outcome. More specifically, the objectives of the study are structured into four main axes:

- Implementation and comparison of classification models: apply and compare the performance of reference algorithms, namely Random Forest, Lasso regularization regression, and Support Vector Machines, in the task of classifying patients depending on whether they were applied the block technique or not.
- valid variables for the training of our different algorithms is a fundamental section in our project and without which it would be impossible for us to advance towards optimal and robust results. For this reason, it will be sought to systematically establish a series of steps with which to process the data in such a way that problems such as their cleaning, the treatment of missing values, the encoding of categorical variables, and normalization are solved.
- Identification and comparison of relevant variables: Different strategies of variable importance extraction will be used to determine which clinical, demographic, and postoperative characteristics are more influential in predicting the success of the anesthetic block.
- Reflection on the strengths and technical limitations: Critically analyze the behavior of the models in such a reduced dataset, addressing the problem of overfitting, the need for regularization techniques, and the implications of transferring these methods to everyday clinical practice.

Taken together, these objectives pursue not only the validation of the methodological approach, but also the demonstration of the potential of machine learning as a support tool for clinical analysis.

3 Dataset

The dataset used in this study was composed of 181 clinical records corresponding to patients who underwent hip or spine surgical procedures in which, for their postoperative period, regional anesthesia was used in half of the cases. Each record initially included a total of 62 variables, which covered a heterogeneous spectrum of information:

- such as age, sex, body mass index, and baseline characteristics.
- such as the type of technique used, the type of anesthesia applied, or the ease of the procedure.
- use of anesthetics, need for pain medication, presence of nausea, or satisfaction with the procedure.

The target variable on which we worked was the use or non-use of anesthetic block in the postoperative period. In this way, we can use libraries such as feature importances with the objective of seeing which variables take greater importance depending on whether one technique or another was applied.

3.1 Data processing

Before applying the models, a preprocessing and cleaning process of the data was carried out, aimed at guaranteeing the consistency and analytical utility of the dataset. The main stages were:

- Screening of variables: The first step was the cleaning of variables with no interest for our model and the creation of new significant variables from the existing ones.
- Cleaning and quality control: Inconsistencies in records were identified and corrected, duplicates were removed, and formats were homogenized.
- Handling of missing values: Simple imputation methods, in our case the mean, were applied in variables with a low percentage of absence; in cases with a greater proportion of missing values, the variables were discarded to avoid biases.
- Encoding of categorical variables: The one-hot encoding technique was used to transform nominal variables into numerical representations compatible with the classification algorithms.
- Normalization, scaling, and elimination of correlated variables: Taking into account that some models can be sensitive to the magnitude of the predictors, the continuous variables were scaled using standard normalization techniques.

This preprocessing flow made it possible to transform a clinical dataset into a consistent analytical matrix, consisting of 180 entries and 54 variables each, suitable for the implementation of the selected machine learning models, with the noteworthy fact that the variables are very little correlated with each other, as can be seen in figure 1, something that may complicate model training in the future since it will make pattern detection more difficult.

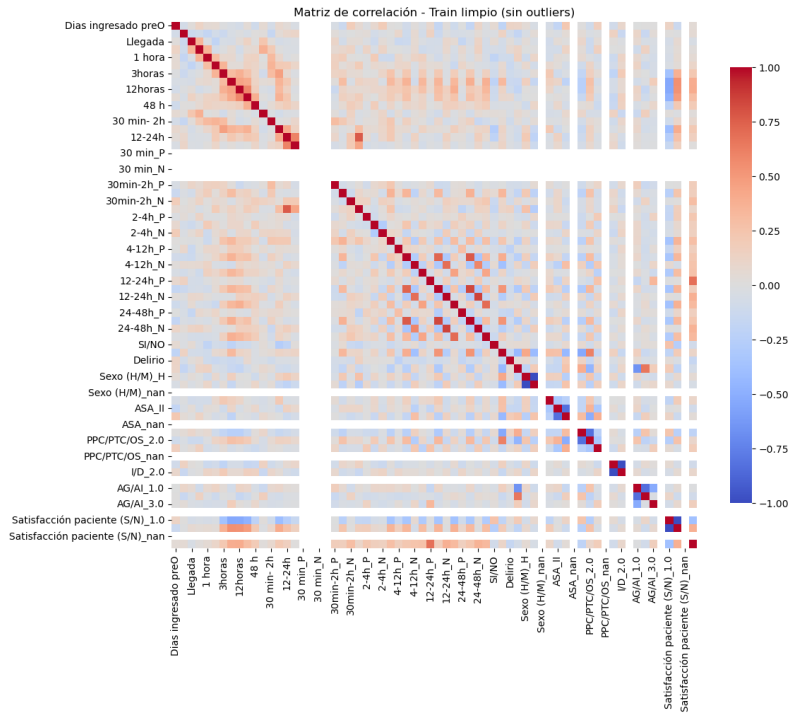


Figure 1: Correlation matrix of the variables before applying high-correlation filtering.

4 Methods

The study was structured as a supervised binary classification problem, with the dependent variable being the use of anesthetic block during surgery (use vs. non-use). Fifty-four clinical, demographic, and postoperative variables were used as predictors, following the cleaning and preprocessing steps described above.

4.1 Algorithms used

Three representative classification algorithms were employed:

- Random Forest (RF): A decision tree ensemble algorithm robust to noise and capable of handling mixed data (categorical and numerical). Also used to estimate variable importance via the Gini criterion or permutation.
- Lasso Regression (L1 penalization): Applied to a logistic model to select relevant predictors by shrinking irrelevant coefficients to zero, reducing dataset dimensionality relative to sample size.
- Support Vector Machine (SVM): Implemented with linear and radial basis function (RBF) kernels to explore linear and nonlinear relationships. Numerical variables were normalized due to the algorithm's sensitivity to scale.

In all models, hyperparameter optimization techniques were applied, including the use of tools such as the GridSearchCV library.

4.2 Experimental design

In order to evaluate model performance and ensure comparability among them, a validation scheme and a set of standardized metrics were defined. The dataset was divided into training and test subsets, applying cross-validation techniques to mitigate the risk of overfitting due to the small sample size.

The selected performance metrics included global classification indicators, such as accuracy, recall, precision, and F1 score. This procedure allowed the establishment of a uniform comparison framework among algorithms, ensuring that the observed performance differences were due to the intrinsic behavior of each model and not to variations in the evaluation protocol.

5 Results

First, it is important to note that the results should be interpreted with caution due to certain limitations of the dataset. On one hand, the records come from a wide age range, which introduces heterogeneity and potential bias in the distribution of the variables. On the other hand, the data encompass three different types of surgical interventions, which may also affect the consistency of the patterns identified and, consequently, the performance of the models.

Training the three proposed models achieved accuracy values above fifty percent in all cases, indicating acceptable performance given the size and characteristics of the available dataset. Although accuracy levels varied between algorithms, and despite the variety of hyperparameters used, the results did not improve. This reinforces the notion of the need to synthetically increase the size of our dataset while also separating it according to the type of surgical procedure performed.

Table 1: Evaluation metrics of the trained models

Models	Accuracy	Precision	Recall	F1-score
Random Forest	0.5278	0.5238	0.6111	0.5641
Lasso L1	0.528	0.524	0.611	0.564
SVM	0.556	0.550	0.611	0.579

In addition, specific techniques were applied to estimate variable importance. These analyses showed that, in cases where the postoperative block technique was used, medication use decreased on average by twenty percent according to the pretrained models. Furthermore, according to the Lasso L1 model, those who did not receive the postoperative block have an eighty percent higher likelihood of reaching the pain threshold variable is, of requiring a much stronger pain analgesic compared to those who did receive the block.

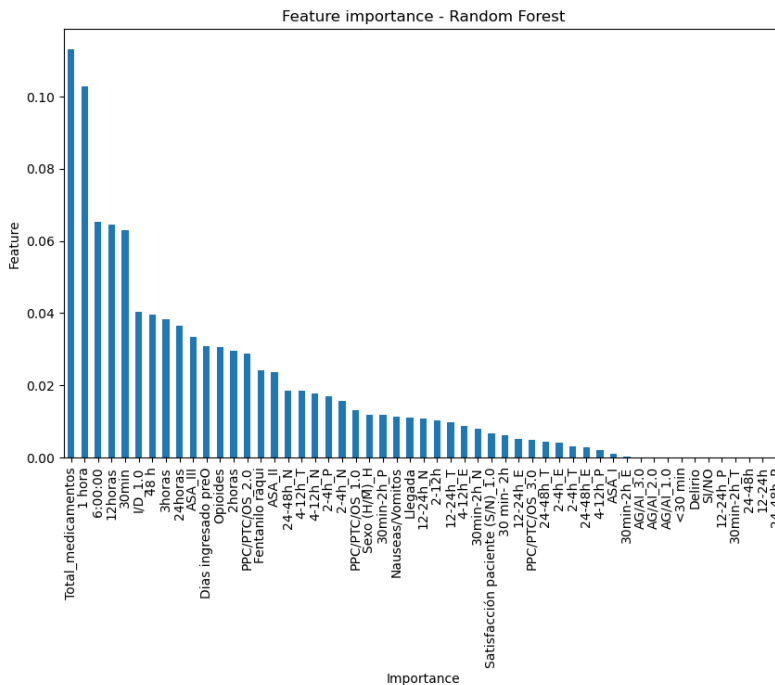


Figure 2: Variable importance after training the Random Forest model

These data are interesting for continuing this research, as, as can be seen in figure 2, the variables with the greatest importance tend to be those related to the use of medications for pain management. This indicates that, even when working with a dataset that appears to have certain biases, we are on the right track. In the future, it would be interesting to explore the possibility of increasing the dataset size, either by adding new data or generating additional data from synthetic datasets.

6 Conclusions

This work demonstrates the viability of applying supervised machine learning techniques to a real clinical dataset with the aim of analyzing the factors associated with the use of anesthetic block in hip and spine surgeries.

First, the implementation and comparison of different classification models, such as Random Forest, Lasso regression, and Support Vector Machines, allowed us to show that, even in a scenario with a reduced number of observations, results can begin to be obtained.

Second, variable importance extraction techniques facilitated the identification of clinical and demographic predictors with the greatest influence on the decision to apply a block, highlighting the potential of these methods not only as predictive tools but also as instruments for interpretation.

Finally, the critical analysis of the behavior of the models highlights both the strengths and the limitations inherent to the adopted approach. Although the obtained results cannot be considered satisfactory, since the heterogeneity derived from including different types of surgical interventions limits the external validity of the conclusions and makes their generalization difficult, they can nevertheless be seen as a promising starting point for future research.

7 Future Research

For future research, it would be of interest to independently analyze each of the surgical procedures included in the present study in order to assess whether the factors associated with the use of anesthetic block differ according to the type of intervention. Furthermore, given the limited size of the dataset, a promising line of work lies in the generation of synthetic data through neural network-based techniques, which would allow for expanding the available sample, enhancing the robustness of the models, and facilitating the validation of findings in broader clinical contexts.

8 Acknowledgements

This work was supported by Universidade da Coruña, Xunta de Galicia and CITIC. TALIONIS Research Group of the University of A Coruña (grants for the consolidation and structuring of competitive research units ED431B 2025/23). CITIC is a center accredited for excellence within the Galician University System and a member of the CIGUS Network (Ref. ED431G2023/01). Additionally, it is co-financed by the European Union through the FEDER Galicia 2021–2027 operational program

Bibliography

- Y.-Y. Li, J.-J. Wang, S.-H. Huang, C.-L. Kuo, J.-Y. Chen, C.-F. Liu, and C.-C. Chu. Implementation of a machine learning application in preoperative risk assessment for hip repair surgery. *BMC Anesthesiology*, 22(1):116, 2022. <https://doi.org/10.1186/s12871-022-01648-y>.
- S. Silvey and J. Liu. Sample size requirements for popular classification algorithms in tabular clinical data: Empirical study. *Journal of Medical Internet Research*, 26(1), 2024. <https://doi.org/10.2196/60231>.
- C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25, 2007. <https://doi.org/10.1186/1471-2105-8-25>.
- W. Wang and T.-W. Pai. Enhancing small tabular clinical trial dataset through hybrid data augmentation: Combining smote and wgan-gp. *Data*, 8(9):135, 2023. <https://doi.org/10.3390/data8090135>.
- C. Wies, R. Miltenberger, G. Grieser, and A. Jahn-Eimermacher. Exploring the variable importance in random forests under correlations: A general concept applied to donor organ quality in post-transplant survival. *BMC Medical Research Methodology*, 23(1):209, 2023. <https://doi.org/10.1186/s12874-023-02023-2>.
- Y. Zhou, S. Wang, Z. Wu, W. Chen, D. Yang, C. Chen, G. Zhao, and Q. Hong. An explainable and supervised machine learning model for prediction of red blood cell transfusion in patients during hip fracture surgery. *BMC Anesthesiology*, 24(1):467, 2024. <https://doi.org/10.1186/s12871-024-02832-y>.