



Rule-based decision tree for surgical prioritization in otolaryngology

Árbol de decisión basado en reglas para la priorización quirúrgica en otorrinolaringología

Fabián Silva-Aravena¹ , Jenny Morales¹ 

¹ Universidad Católica del Maule. Facultad de Ciencias Sociales y Económicas. Talca, Chile.

DOI: 10.64966/ingeniare.v34.07

Abstract

Effective waitlist management for surgery in otolaryngology (ENT), as in most public health care programs, remains a significant issue. In the present study, using the Classification and Regression Tree (CART) algorithm, a decision tree model based on biopsychosocial factors was developed to classify 205 patients into three priority classes (High, Medium, and Low) based on a continuous score assigned by the attending physicians. The last tree considered severity (Sev), urgency (Urg), and being able to attend the study (Dest) as the main predictors, yielding five validated clinical rules. With an overall accuracy of 82,3%, the model shows its predictive power and practical relevance. Its transparency and clinical reasoning underpin its trustworthiness as a decision aid. The next steps will include interfacing with dynamic prioritization methods.

Keywords: Surgical Prioritization, decision tree, clinical explainability, otolaryngology, CART algorithm.

Resumen

La gestión eficaz de las listas de espera para cirugía en otorrinolaringología, al igual que en la mayoría de los programas de salud pública, sigue siendo un problema importante. En el presente estudio, utilizando el algoritmo de árbol de clasificación y regresión (CART), se desarrolló un modelo de árbol de decisión basado en factores biopsicosociales para categorizar a 205 pacientes en tres clases de prioridad (Alta, Media y Baja) a partir de una puntuación continua asignada por los médicos tratantes. El último árbol consideró la severidad, la urgencia y la dificultad para estudiar como principales predictores, generando cinco reglas clínicas validadas. Con una precisión general del 82,3 %, el modelo demuestra su poder predictivo y relevancia práctica. Su transparencia y razonamiento clínico respaldan su fiabilidad como herramienta de ayuda para la toma de decisiones. Los próximos pasos incluirán la interacción con métodos dinámicos de priorización.

Palabras clave: Priorización quirúrgica, árbol de decisión, explicabilidad clínica, otorrinolaringología, algoritmo CART.

INTRODUCTION

Elective surgery waiting lists are an important operational and ethical challenge for global public health systems [1], [2]. In many middle- and low-income countries, such as Latin American countries, limited surgical capacity, resource scarcity, and a growing population have led to long waiting times for non-emergent operations [3]. This is particularly relevant in subspecialties, such as ENT, where many diseases do not affect a patient's life on a life-and-death basis but do affect quality of life, communication, social integration and mental health functioning [4] - [6].

Despite numerous policy interventions, the scheduling of surgeries is still frequently based on simple mechanisms, such as First-Come-First-Served (FCFS), that ignore patients' multidimensional vulnerability [7]. The most recent literature has highlighted the need to develop more equitable and clinically based prioritization models. Numerous methods have been proposed, from composite score systems to sophisticated machine learning (ML) [6] - [12]. For example, some models combine clinical severity with socioeconomic factors to more evenly distribute surgery slots [13]. Others use predictive analytics to predict the risk of deterioration or procedural success. However, these approaches have limited transparency and interpretability (that is, black-box models) and are also difficult to integrate into real clinical workflows. Furthermore, models designed solely for predictive performance also give little or no consideration to physician trust, institutional fit, and practicality of their use, all critical components for the effectiveness of any decision support tool in the clinical setting [14], [15].

In the current climate, easily interpretable models, such as decision trees, have seen a resurgence, as they can help translate data science to clinical practice [16], [17]. For example, the CART algorithm can generate a simple rule-based model that is mathematically accurate and intuitive to medical doctors [18], [19]. Properly applied, CART models can be a representation of clinical reasoning, provide actionable insights, and function as an interface between algorithmic output and human intuition. That makes them particularly applicable to surgery scheduling committees, where high transparency, reasonable selection criteria, and justifiable pathways are needed.

The primary objective of this study is to create and interpret a clinically useful decision tree model that can be used to aid in the preoperative process for elective patients with ENT [20], [21]. 205 patients were collected from the high-complexity public hospital (HCPH), characterized by biopsychosocial factors and a research-scaled prioritization score (Pscore) assigned by a clinical team. A CART model was trained to predict priority levels. The model structure was then operationalized into a set of rules, which was reviewed by medical experts to ensure agreement with hospital policies and clinical practice. The core model variables (severity, urgency, and capacity to study) were determined using statistical methods and expert judgment.

The central research question is: Can a clinically validated, rule-based decision tree based on biopsychosocial variables accurately and transparently classify ENT surgical patients into priority levels? the hypothesis is that such a model, if based on key clinical dimensions, can achieve sufficient predictive accuracy while preserving interpretability for use in real-world surgical scheduling.

The specific contribution of this study is that a clinical data-validated rule-based model can provide a practical, transparent, and implementable solution to the problem of surgical prioritiza-

tion if relevant decision-makers contribute to the validation. Unlike black-box models, the tree is directly interpretable, challengeable, and modifiable by clinical teams. Importantly, its simplicity makes it potentially suitable for low-resource settings lacking in digital infrastructure but needing fair and effective prioritization. This contribution highlights the importance of aligning statistical validation with clinical explainability. It was demonstrated that a CART-based model, rigorously trained and expert-validated, can bridge the gap between predictive performance and real-world decision-making, offering a scientifically grounded, transparent solution for surgical prioritization.

LITERATURE REVIEW

The problem of the surgical waiting list has been analyzed in detail regarding its operational, ethical, and clinical impact [21]. Traditional priority systems, including FCFS and urgency-based ordering systems, have been commonly used in public health facilities but are limited in their ability to accommodate diverse dimensions of patient vulnerability [7]. These methodologies often do not consider psychosocial, functional, or contextual factors that can predict patient outcomes on extensive waiting lists. Consequently, there has been growing attention to holistic approaches that consider a variety of biopsychosocial indicators to support scheduling decisions [20].

Some composite prioritization indices have been proposed in the literature to fill this void. For example, the New Zealand Priority Criteria Project and the Western Canada Waitlist Project proposed a multiattribute concept that includes clinical severity, symptom intensity, function impact, and social indicators [22], [23]. These indices, although based on reasonable expert consensus, are often static, use arbitrary scoring systems, and may not be sensitive to institutional changes over time.

Neurologically based survival prediction models were also used in contexts such as Parkinson's disease [24], but subsequently replaced by a variety of more data-driven approaches over the last decade, including logistic regression, random forests, and neural networks to predict risk of decompensation or likelihood of delisting after waiting list [25]. Although such models can achieve high predictive power, their complexity hinders interpretability and clinical adoption.

For interpretability, some authors seek a rule-based model, namely, decision trees. CART and its successor C4.5 score have been used to classify patients according to clinically oriented thresholds [26]. These models provide transparent and easy-to-follow decision paths that align with clinical reasoning, making them especially valuable for settings where physician trust and explainability are essential. However, such efforts are computationally demanding, and practitioners still have a black-box view of these experimental approaches [27]. Furthermore, the majority of the models focus on the performance itself, rather than the user practicality, and therefore are not feasible for day-to-day usage in the hospital.

Within the field of ENT, there is a striking absence of specialty-specific prioritization instruments. Despite the presence of common models for elective surgery, there is limited literature on the special aspects of ENT pathologies (that is, modifications in communication, mild chronic pain, social life) in terms of health-related quality of life. Moreover, there are few descriptions on

enhancing model-based decisions with clinical judgment in low-resource settings, where model accountability and implementation feasibility are crucial.

This article distinguishes itself from previous literature by providing a clinically interpretable decision tree, based on real patient data and consistent with expert judgment, for prioritization in otorhinolaryngology. Unlike approaches using manually crafted rules or black-box models presented in the literature, the method learns the decision rules used to classify examples by training a CART model on a data set based on the vulnerability of patients on the surgical waiting list. The rules are then checked for clinical consistency, ensuring a balance between their rigor and their utility in clinical practice. This hybrid approach (data-driven but explainable) contributes to clinical decision-making.

METHODOLOGY

In this paper, an explanatory and classificatory model is proposed to prioritize surgical assistance for patients in the ENT service of an HCPH in Chile. The model included clinical, social, and psychological factors. The CART methods were used because they are capable of producing reproducible, transparent decision rules that can be used as a clinical decision-making aid.

Study design and data set

An observational, retrospective, and quantitative study was conducted on anonymized data from $N = 205$ patients waiting for ENT surgery. Each patient was described using 20 independent variables, $x_1, x_2, \dots, x_{20}$, corresponding to the clinical and psychosocial profiles. In addition, an ordinal Pscore value assigned a priori by the ENT clinical team was considered. Factors included clinical (severity, urgency, comorbidity, pain), psychosocial (ability to work, transferring inconvenience, patient care for another person), and other variables (e.g., type of residence area, among others). This information was validated and authorized by the Scientific Ethics Committee of the Universidad Católica del Maule in Evaluation Report No. 09/2025.

Target variable: Pscore

The Pscore is a continuous standing of biopsychosocial susceptibility established by the ENT clinical team. This score combines medical, social, and functional measures and ranges from the Interval 0,1, with values close to 1 reflecting the increasing need for the patient to undergo surgery. It was found that for training the classifier, an ordinal categorical variable was more appropriate for this continuous indicator. This transformation was performed to allow metalearning multiclass decision tree construction while not obliterating the hierarchical structure, which is naturally contained in the Pscore. This new variable Y_i is called: $Y_i \in \{Low, Medium, High\}$, where the categories were determined by the terciles of the data. For the former, it was the 95th percentile of the treatment probability ($Q_1 \sim 0.33$ and $Q_2 \sim 0.66$). The transformation rule used was as follows:

$$Y_i = \begin{cases} \text{Low, if } Pscore_i < Q_1 \\ \text{Medium, if } Q_1 \leq Pscore_i < Q_2 \\ \text{High, if } Pscore_i \geq Q_2 \end{cases}$$

This discretization method was designed for several methodological reasons. One is that it allows us to deal with balanced classes, which is crucial to prevent bias in the supervised learning models [28], [29]. Maintain the ordinality of logic because each of the lower, middle, and higher levels respects the order of the Pscore. Lastly, the model transformation set (the model before and after the transformation is the same) improves clinical interpretability by providing classes that are sensible for both medical and administrative groups working on surgical prioritization. The last result was a fairly equal distribution of patients in the smoothed categories: one class considered as “low priority” with 69 observations, one class of “medium priority” with 68 observations, and one class for “high priority” with 68 observations. The new variant categorical variable Y_i was applied as a dependent variable in the CART decision tree.

Table 1 describes the complete set of predictor variables considered during model building, along with their corresponding clinical or social descriptions, data types, and allowable values. Indicated with a (*) are the variables that did not have enough variance and were excluded during preprocessing.

Table 1. Description of predictor variables considered in the CART model.

Variable	Description	Values / Categories
Sever	Severity	1-3 [low, medium, high]
Urg	Urgency level	1-10
Jclin	Maximum waiting time	0-10 [30-day interval, e.g., 1-30 days = 1, 31-60 days = 2, and so on]
Tsuen	Sleep disorder	1-3 [low, medium, severe]
Tlist*	Time on surgical waitlist	1-10 [30-day interval, e.g., 1-30 days = 1, 31-60 days = 2, and so on]
Pmxc	Expected improvement due to surgery	1-3 [low, medium, high]
Dest	Capacity to study	1-3 [NA, no, yes]
Com	Chances of developing comorbidities	1-3 [low, medium, high]
Lfam	Capacity of participating in family activities	1-3 [NA, no, yes]
Hanor	Affected area	1-3 [no, low presence, high presence]
Opat	Presence of other pathologies	1-2 [no, yes]
Diag	Diagnosis	1-20 [diagnostic 1, ..., diagnostic 20]
Olim	Other limitations	1-3 [no, medium, severe]
Ncuid	Needs of a caregiver	1-2 [no, yes]
Rcuid	Patients cares for another person	1-2 [no, yes]
Dolor	Pain scale	1-10
Dtrab	Capacity to work	1-3 [NA, no, yes]
Acc	Type of residence area	1-3 [urban, rural, high rurality]

Table 1. Description of predictor variables considered in the CART model. (continued)

Variable	Description	Values / Categories
Dtras	Difficulty in transferring	1-2 [no, yes]
Ccrit*	Need for clinical bed	1-2 [no, yes]

Note. Variables marked with (*) were removed due to low variance.

Predictor variables and preparation

All original variables available for the dataset were included in the construction of the CART model as predictors, with two exceptions: the Pscore (which was used only to create a categorical target variable as previously described) was not included as a predictor. A total of twenty predictors of biopsychosocial variables were included (Table 1). Those covariates describe an in-depth biopsychosocial framework of the study, to cover the different dimensions that could have a bearing on the time required for surgical intervention. Before including these variables in the model, a systematic testing and a preprocessing scheme were performed based on the following principles.

Data completeness

First, the missing values in the raw data set were examined. A manual check confirmed that there were no missing or empty values in any of the twenty predictors. The prevalence of the lack of missing data was ascribed in part to the systematic and comprehensive nature of database construction carried out by the clinical team, which reinforces the reliability of the analysis and informs the potential for model replicability in a real-world clinical setting.

Explanatory variance

Second, the extent to which each predictor variable can vary was tested, with the aim of removing those with almost constant values. In particular, the percentage of the most common category was examined in each variable, and a cutoff point of 95% was set to discard it. In other words, if less than 95% of the patients had the same value, the low variance was considered and removed from the model. Two variables have been selected that were not considered for the prediction using this rule. The first was Tlist (waiting list time), with the highest category being 97% of the sample. The second was Ccrit (the requirement for critical care beds), with 99% of patients belonging to the same class. Since they had low diversity, both also carried little information about the discriminatory capacity of the CART model and brought potential noise or redundancy to the classification process, so it was decided to remove them.

Clinical and social relevance

The third standard used was the theoretical and empirical relevance of the variables, based on the previous scientific literature and the expert opinion of the otolaryngologists who participated in the study. This estimate ensured that the included factors were not only statistically significant, but also consistent with the real-life determinants that influence the need for surgical treatment from a medical and social perspective.

Coding and transformation

Given that the CART algorithm can handle both categorical and continuous predictors, no data processing method (e.g., one-hot encoding or normalization) was performed on the original encoding of ordinal and nominal predictors [30]. This method allows the model to automatically learn the optimal cut-off points for ordinal variables, thereby generating transparent, reproducible decision trees and extracting interpretable rules, as proposed by the study's explanatory purposes. The last variables included in the model were divided equally between clinical and social content to cover all potential determinants of surgical prioritization in the study patient population.

Cart decision tree model

Mathematical foundation

CART is a popular supervised learning algorithm used to solve both regression and classification problems [30] - [31]. In the analysis, the multiclass classification form of the CART algorithm was used to predict the degree of surgical priority ($Y_i \in \{Low, Medium, High\}$) given a vector of biopsychosocial predictor variables.

In other words, the algorithm aims to partition $x \subseteq \mathbb{R}^p$ into a finite number of regions $\{R_1, R_2, \dots, R_M\}$ where the class ratio or quality within a region is similar across all regions. Each region R_m consists of a finite sequence of these decision rules in the form:

$$x_j \leq c \text{ o } x_j > c$$

where x_j is a feature and c is an optimal threshold learned during training.

Then, the criterion used to assess the quality of the partition is the Gini index, which measures the node's impurity t [32], [33]. Let $\hat{p}_j(t)$ represent the proportion of observations belonging to the class j at node t , and K is the total number of classes ($K = 3$ in this specific problem). The Gini impurity is consequently given by equation (1):

$$G(t) = 1 - \sum_{j=1}^K \hat{p}_j(t)^2 \quad (1)$$

When the node has mainly samples from one class, this value becomes minimum.

a tree by recursively selecting the best split at each step was built: At each step, the split selected results in the lowest weighted impurity of the newly created child nodes, equation (2):

$$\Delta G = G(t) - \left(\frac{n_L}{n_t} G(t_L) + \frac{n_R}{n_t} G(t_R) \right) \quad (2)$$

where t_L and t_R are the left and right children after the split, and n_t , n_L , and n_R are the number of observations in the node, left child node, and right child node, respectively.

This partitioning process was repeated recursively until a stopping criterion was reached, for example, a tree with maximum depth or a terminal node containing the fewest observations.

MODEL TRAINING AND VALIDATION

The dataset was split into a 70% training set and a 30% test set to evaluate the model's predictive ability and generalizability; 10-fold cross-validation was also applied, and the sample size was not particularly large (only 205 cases). This reduces the risk of overfitting and yields more stable, reliable performance estimates when segmenting patient data.

Some parameters used to curb complexity and ensure interpretability in the model are:

- Impurity measure: Gini index (the criterion used to select the best splits).
- Maximum depth: 4.0, with up to four multi-way splits for interpretation and simplification of tree branches.
- Minimum terminal node: 10 persons, to ensure a minimum amount of support at the leaves.
- Pruning criteria: Prune the tree with cost-complexity pruning and use 10-group cross-validation. The best subtree was chosen at one standard error from the minimum cross-validation error (α -pruning).

The analysis was implemented in Python 3.12 using the scikit-learn library, a simple-to-use yet extensive tool for decision tree classifiers.

Final variables selection

The model's clinical interpretability was improved without sacrificing its predictive performance; thus, a model selection procedure was performed based on theoretical and empirical background. This process allowed us to subselect the characteristics of the entire set of predictors $X = \{x_1, x_2, \dots, x_p\}$ into an optimal subset $X^* \subset X$, retaining only influential, clinically validated, and independently prognostic factors.

Mathematically, the importance of each variable of $x_j \in X$ was measured by its cumulative contributions to decrease the value of the Gini impurity index in the decision tree growth process in equation (2). The cumulative importance of variable x_j was calculated in equation (3):

$$Imp(x_j) = \sum_{t \in T_{x_j}} \Delta G(t, s_j) \quad (3)$$

where T_{x_j} is the collection of nodes in which x_j was used to split. Additionally, the model's sensitivity was assessed by sequentially leaving out each predictor using a leave-one-variable-out cross-validation strategy. It was specifically considered Acc_X the accuracy of the model built with all predictors, and $Acc_{X \setminus x_j}$ the accuracy without the variable x_j , and the relative loss of accuracy was estimated as, equation (4):

$$\Delta Acc(x_j) = Acc_X - Acc_{X \setminus x_j} \quad (4)$$

These two statistics, Gini importance and exclusion accuracy loss, allowed us to select a reduced subset X^* that maximized the model's explanatory power and predictive stability [34]. Finally, this selection was clinically validated with the hospital's medical personnel to ensure that the included variables were reasonable, relevant, and applicable in the context of actual surgical prioritization.

RESULTS

Importance of variables and cart model results

Table 2 provides an interpretation of the results to better understand the model, the reasons it works, and explains its internal logic, with particular attention to the variable Dest. This variable has a Gini importance of 0.17, indicating it contributed fairly well to reducing node impurity during decision tree construction. However, when “Dest” was omitted from the model, the percent decrease in overall accuracy reached 6,2%, the highest among all factors.

This seeming paradox reveals a key observation: While “Dest” was sparse in the splits closest to the root, it was highly relevant at certain parts of the tree for which its presence increased classification performance. It is also worth mentioning that even though its global Gini importance is lower than for “Severity” and “Urgency,” it shows high conditional importance through multiple interactions.

In contrast, the variable Severity, with the highest Gini importance (0.32), led to a reduction in elimination precision of only 0.4%. This indicates that “Severity” was always considered in the first splits, which is useful to the tree structure.

Finally, it was found that Hanor and Dtrab had the lowest Gini importance values (0.11 and 0.10, respectively) and the least accuracy decrease when removed (1.0% and 0.5%, respectively). This would indicate that even if these predictors may have some contextual relevance in particular cases, their statistical contribution was limited in the sample. Thus, the choice not to include them in the final decision tree structure is justified by clinical simplicity, interpretability, and performance metrics.

Table 2. Gini importance and accuracy loss for key variables in the CART model.

Variable	Gini Importance	Accuracy Loss by Omission (%)
Severidad	0,32	0,4
Urgencia	0,3	3,3
Dest	0,17	6,2
Hanor	0,11	1
Dtrab	0,1	0,5

Evaluation metrics

An overall precision of 82,3% was obtained, meaning that more than four out of five patients had their priority level correctly predicted by the model. Moreover, the balanced accuracy was obtained, which is an average of the sensitivity in different classes and was found to be 82,1%. This is an indication of a model that is strong in dealing with balanced class distributions.

As an auxiliary analysis, the confusion matrix was generated. As illustrated in [Figure 1](#), this visualization helped investigate the types of misclassification introduced across priority levels. The model occasionally misclassifies between the 'Medium' and 'High' groups, and such confusion between intermediate conditions is clinically reasonable when these conditions tend to share common characteristics.

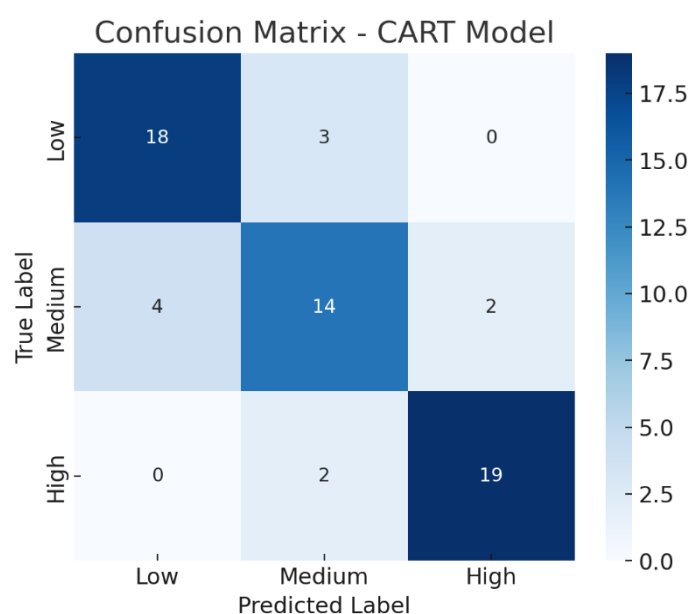


Figure 1. Confusion matrix for the CART model predictions across priority levels.

In addition to overall accuracy (82,3%) and balanced accuracy (82,1%), the precision, recall, and F1 score for each class are presented to provide a more comprehensive assessment of model performance ([Table 3](#)). The model performs best in the “High” and “Low” categories, with F1 scores of 0,90 and 0,84, respectively. The middle class, meanwhile, shows slightly lower differences ($F1 = 0,72$), consistent with the overlap observed in the confusion matrix and reflecting the clinical ambiguity of intermediate cases. These findings underscore the model's robustness while also pointing to areas for future improvement.

Table 3. Evaluation metrics for each priority class.

	Precision	Recall	F1-Score
Low	0,82	0,86	0,84
Medium	0,74	0,7	0,72
High	0,9	0,9	0,9
Accuracy	0,823	0,823	0,823
Macro avg	0,82	0,82	0,82
Weighted avg	0,82	0,82	0,82

These metrics, along with with decision tree-based models from surgical prioritization, are also compared and presented in recent literature. The model demonstrates competitive performance while maintaining high interpretability, an essential aspect for its practical implementation in public health systems.

This set of metrics confirms that the model is not only statistically robust but also operationally valuable, providing a reliable foundation to support decision-making in real-world clinical settings.

Interpretability and clinical validation

One of the main advantages of the CART model lies in its high degree of explainability. The model was trained using predefined hyperparameters, and the resulting structure is shown in [Figure 2](#). From this tree, five main decision rules were extracted, representing clinically meaningful patterns. These rules were reviewed by medical professionals to ensure their alignment with expert judgment and real-world practice.

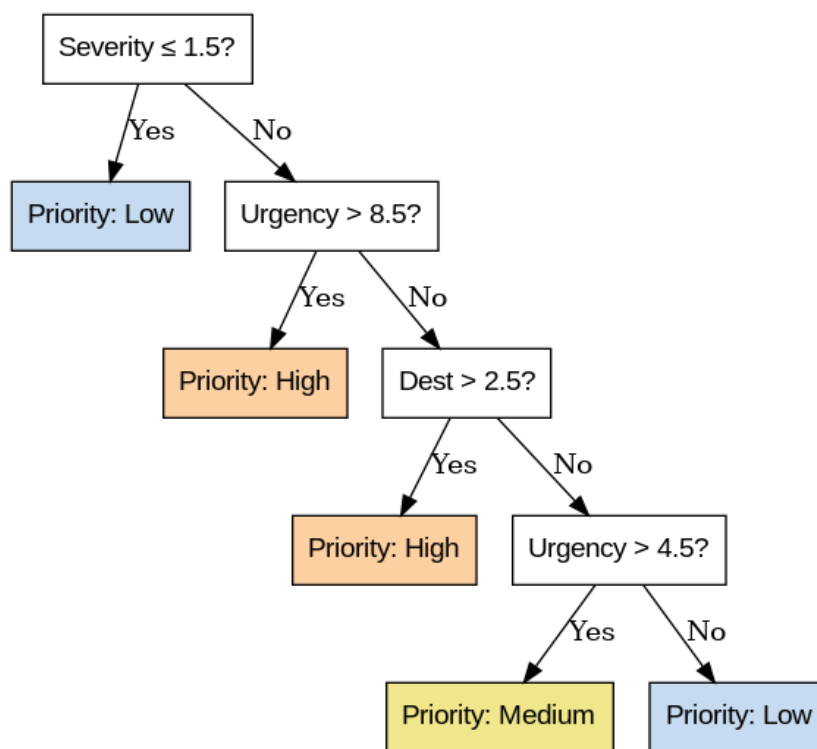


Figure 2. Simplified CART decision tree showing surgical priority based on Severity, Urgency, and Dest.

The most clinically relevant decision rules derived from the model are now presented:

Rule 1: If *Severity* ≤ 1.5 → Low Priority

Rule 2: If *Severity* > 1.5 and *Urgency* > 8.5 → High Priority

Rule 3: If *Severity* > 1.5 and *Urgency* ≤ 8.5 and *Dest* > 2.5 → High Priority

Rule 4: If *Severity* > 1.5 and *Urgency* ≤ 8.5 and *Dest* ≤ 2.5 and *Urgency* > 4.5 → Medium Priority

Rule 5: If *Severity* > 1.5 and *Urgency* ≤ 8.5 and *Dest* ≤ 2.5 and *Urgency* ≤ 4.5 → Low Priority

This level of interpretability not only supports analytical transparency but also serves as a practical decision aid for clinical environments, enhancing traceability and trust in surgical prioritization decisions.

DISCUSSION

In the present study, a clinically interpretable decision tree model was developed to prioritize surgery allocation for elective ENT patients. Unlike purely data-driven training, the approach was based on a group of predefined clinically validated rules derived from expert consensus. The model was constructed on three central concepts: Severity, Urgency, and Dest, which articulate the core aspects of surgical need, psychosocial vulnerability, and expected discharge. The

resulting decision tree proved to be extremely simple, self-evident, and practical to translate into guidelines for hospital practice staff.

The model has good interpretability, and that is one of its main advantages. Although more advanced ML models like random forest or neural network may lose interpretability, the CART model will be understandable from the clinical point of view, so that the clinical teams can understand the decision paths and validate or reject each rule. In particular, the root node is Severity, a concept relevant to the medical context, as it represents the severity of the pathology and aligns perfectly with the way medical experts build their reasoning, readily conveying face validity to the model. In addition, the model also showed good discrimination in classifying patients into low, medium, and high-priority groups, thereby increasing its credibility and reliability in clinical application.

The approach is limited in its predictive flexibility because the model has rule-based rigidity built into its design. The tree models may not be able to detect any hidden patterns not present in the data or changes of patients' status over time. Furthermore, the prioritization mechanism is based on a discretization of the Pscore index that, although it has a clinical basis, imposes arbitrary cut-off values that may not fully capture the continuous risk scales. Hybrid models that integrate expert rules as priors with more flexible learning frameworks should be further explored.

Implementation-wise, there are a number of concerns that need to be addressed to make these practical for use. The potential solution must integrate with existing clinical practice and not impose an additional burden on staff or harm patient equity. For this transport, it is suggested that the decision tree system be implemented within a multidisciplinary scheduling committee as part of a pilot study, given that its transparency can promote discussion and consensus. Additionally, regular model reevaluations are needed to ensure that decision output aligns with emerging clinical practice or resource supply.

While decision trees are interpretable and suitable for diverse datasets, they may be less scalable when dealing with complex or high-dimensional data. It is acknowledged this limitation clarifies that the model was specifically designed for a clinical dataset of 205 observations and 20 structured features. It is believed that for large-scale implementation, the use of hybrid models (e.g., ensemble methods or model stacking) with regular retraining could ensure robustness and maintain explainability.

The study presents some limitations. The number of patients was enough to explore the problem, but it may not fully represent complex cases. Since a rule-based model was used, it might miss more complex patterns in the data. Also, grouping the prioritization scores into three groups, even with expert input, adds some subjectivity. Lastly, because the data came from a single public hospital, the model might not work the same way in other settings. These are all important points for future work, in which the model is planned to be tested more widely and more flexible methods explored.

In conclusion, the rule-based decision tree provides a robust, interpretable, and clinically plausible solution to surgical waitlist prioritization in ENT. It incorporates commonly available factors and is computationally undemanding; thus, it is readily applicable to impoverished public hospitals. Although additional validation will be required in larger patient populations and across

different specialties, the findings suggest that organized clinical reasoning, cast as algorithms, may provide meaningful assistance in the pursuit of fair and efficient surgical scheduling.

CONCLUSIONS

A clinically interpretable decision tree model was developed and applied to order ENT services according to predefined rules derived from expert consensus. By focusing the model formally on 3 main predictors, severity, urgency, and Dest, a simplistic model was successfully designed, stable enough to rationalize the algorithm to enable transparent prioritization in realistic surgical scheduling settings. The tree structure makes sense and provides immediate interpretability, which is crucial to trust and acceptance among health professionals.

The classifier effectively stratified patients into low, medium, and high-priority groups according to clinically validated criteria. Visual examination of the tree verified that the hierarchy of decision rules is maintained with severity as the primary splitting criterion. This further reinforces the correspondence between algorithmic logic and clinical logic. Furthermore, the simplicity of the model facilitates an implementation that can be easily followed, which is especially attractive under conditions of restricted technical resources and data science knowledge.

The results of this study will be used to improve the model, including patient-level data-driven thresholds and dynamic learning from incoming patients, using ensemble methods to maintain interpretability in future iterations. In addition, adaptive calibration methods could be designed to recalibrate decision thresholds at regular intervals in light of changes in clinical need, capacity, or prioritization rules. These modifications would enable the model to adapt and remain relevant over time, without compromising the model's transparency.

Successful execution of CART requires managing institutional and cultural obstacles. It is essential to involve clinical teams in the initial stages, train them on the deployment and interpretation of the model, and seamlessly integrate the tool into the working schedules. A governance mechanism must be implemented not only to ensure scheduled audits and ethical oversight, but also to resolve disputes between the perception of model results and clinical judgment.

As future directions, the model is intended to be extended to other surgical subspecialties and to be evaluated during a prospective pilot implementation. It is also planned to include patient-reported outcomes and social determinants of health to further enhance the prioritization framework. Ultimately, it is hoped that transparent, clinically validated, and operationally practical models like the one studied can inform more equitable and efficient surgical waitlist management, especially in resource-limited health care systems.

ACKNOWLEDGMENTS

The authors thank the ENT unit of the high-complexity public hospital in Chile for providing the data used in this manuscript. This research was funded by the “ANID Fondecyt Iniciación a la Investigación 2024 N° 11240214”.

REFERENCIAS

- [1] T. Stafinski, F. N. Nagase, M. E. Brindle, J. White, A. Young, S. Beesoon, and D. Menon, “Reducing wait times to surgery—an international review,” *Journal of Hospital Management and Health Policy*, vol. 6, 2022. [Online]. Available: <https://jhmhp.amegroups.org/article/view/7456> [↑]
- [2] C. D. Santana Lemos, A. M. M. Magalhães, D. S. T. dos Reis, A. Y. T. de Andrade, K. C. de Almeida, F. Z. Martins, and V. D. B. Poveda, “Access to healthcare: waiting time until the surgical procedure,” *Annals of Medicine*, vol. 57, no. 1, 2025, <https://doi.org/10.1080/07853890.2025.2452358> [↑]
- [3] C. E. Galle, “Walk-In Telehealth Visits Improve Access to Care, Wait Times, and Patient Satisfaction in Adult Patients with Non-Emergent Medical Conditions,” Ph.D. dissertation, School of Education, Brandman University, Irvine, CA, USA, 2021. [Online]. Available: <https://www.proquest.com/openview/e187ab77831334a783fa35c02f7f869c/1?pq-origsite=gscholar&cbl=18750&diss=y> [↑]
- [4] J. M. Patterson, R. Govender, J. Roe, G. Clunie, J. Murphy, G. Brady, and P. Carding, “COVID-19 and ENT SLT services, workforce and research in the UK: A discussion paper,” *International Journal of Language & Communication Disorders*, vol. 55, no. 5, pp. 806–817, 2020, <https://doi.org/10.1111/1460-6984.12565> [↑]
- [5] A. A. Altamimi, M. Robinson, E. M. Alenezi, J. Kuthubutheen, T. Veselinović, G. Bernabei, and C. G. Brennan-Jones, “The impact of waiting times on behavioral outcomes for children with otitis media,” *Telemedicine Reports*, vol. 4, no. 1, pp. 359–365, 2023, <https://doi.org/10.1089/tmr.2023.0055> [↑]
- [6] F. Silva-Aravena, H. N. Delafuente, and C. A. Astudillo, “A novel strategy to classify chronic patients at risk: A hybrid machine learning approach,” *Mathematics*, vol. 10, no. 17, Art. no. 3053, 2022, <https://doi.org/10.3390/math10173053> [↑]
- [7] S. Youn, H. N. Geismar, and M. Pinedo, “Planning and scheduling in healthcare for better care coordination,” *Production and Operations Management*, vol. 31, no. 12, pp. 4407–4423, 2022, <https://doi.org/10.1111/poms.13867> [↑]
- [8] F. Silva-Aravena, J. H. Gutiérrez-Bahamondes, H. Núñez Delafuente, and R. M. Toledo-Molina, “An intelligent system for patients’ well-being: A multi-criteria decision-making approach,” *Mathematics*, vol. 10, no. 21, Art. no. 3956, 2022, <https://doi.org/10.3390/math10213956> [↑]
- [9] F. Silva-Aravena and J. Morales, “Dynamic surgical waiting list methodology: A networking approach,” *Mathematics*, vol. 10, no. 13, 2022, <https://doi.org/10.3390/math10132307> [↑]
- [10] F. Silva-Aravena, J. Morales, M. Jayabalan, M. E. Rana, and J. H. Gutiérrez-Bahamondes, “Dynamic Surgical Prioritization: A Machine Learning and XAI-Based Strategy,” *Technologies*, vol. 13, no. 2, Art. no. 72, 2025, <https://doi.org/10.3390/technologies13020072> [↑]
- [11] F. Silva-Aravena, J. Morales, and M. Jayabalan, “e-Health strategy for surgical prioritization: a methodology based on Digital Twins and reinforcement learning,” *Bioengineering*, vol. 12, no. 6, Art. no. 605, 2025, <https://doi.org/10.3390/bioengineering12060605> [↑]
- [12] M. Bagheri, M. Bagheritabar, S. Alizadeh, M. S. S. Parizi, P. Matoufinia, and Y. Luo, “Machine-learning-powered information systems: A systematic literature review,” *Applied Sciences*,

- vol. 15, no. 1, Art. no. 296, 2024, <https://doi.org/10.3390/app15010296> [†]
- [13] J. Van De Klundert, H. De Vries, F. Pérez-Galarce, N. Valdes, and F. Simon, “The effectiveness, equity and explainability of health service resource allocation—with applications in kidney transplantation and family planning,” *Frontiers in Health Services*, vol. 5, Art. no. 1545864, 2025, <https://doi.org/10.3389/frhs.2025.1545864> [†]
- [14] N. Jayatilleke and S. Weerakoon, “Machine learning applications in decision support systems for healthcare governance,” *International Journal of Advanced Scientific Computation, Modeling, and Simulation*, vol. 14, no. 10, pp. 1–24, 2024 [†]
- [15] M. M. H. Ennab, “A hybrid convolutional-fuzzy model for interpretable ai in healthcare: improving transparency and accuracy in chronic disease Management,” Ph.D. dissertation, Department of Computer Science and Mathematics, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, 2025 [†]
- [16] C. Giordano, M. Brennan, B. Mohamed, P. Rashidi, F. Modave, and P. Tighe, “Accessing artificial intelligence for clinical decision-making,” *Frontiers in Digital Health*, vol. 3, 2021, <https://doi.org/10.3389/fdgth.2021.645232> [†]
- [17] E. Barbierato and A. Gatti, “The challenges of machine learning: A critical review,” *Electronics*, vol. 13, no. 2, Art. no. 416, 2024, <https://doi.org/10.3390/electronics13020416> [†]
- [18] J. Jung, C. Concannon, R. Shroff, S. Goel, and D. G. Goldstein, “Simple rules to guide expert classifications,” *Journal of the Royal Statistical Society Series A*, vol. 183, no. 3, pp. 771–800, 2020., <https://doi.org/10.1111/rssa.12576> [†]
- [19] J. Hao, S. Luo, and L. Pan, “Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes,” *Scientific Reports*, vol. 12, no. 1, Art. no. 9858, 2022, <https://doi.org/10.1038/s41598-022-14143-8> [†]
- [20] F. Silva-Aravena, E. Álvarez-Miranda, C. A. Astudillo, L. González-Martínez, and J. G. Ledezma, “Patients’ prioritization on surgical waiting lists: A decision support system,” *Mathematics*, vol. 9, no. 10, Art. no. 1097, 2021, <https://doi.org/10.3390/math9101097> [†]
- [21] D. Rathnayake, M. Clarke, and V. Jayasinghe, “Patient prioritisation methods to shorten waiting times for elective surgery,” *PLOS ONE*, vol. 16, no. 8, 2021, <https://doi.org/10.1371/journal.pone.0256578> [†]
- [22] J. Whitehead, “Prioritizing sustainability indicators: Using materiality analysis,” *Business Strategy and the Environment*, vol. 26, no. 3, pp. 399–412, 2017, <https://doi.org/10.1002/bse.1928> [†]
- [23] C. De Coster et al., “The Western Canada Waiting List Project: The Western Canada Waiting List Project: Development of a priority referral score for hip and knee arthroplasty,” *Journal of Evaluation in Clinical Practice*, vol. 13, no. 2, pp. 192–197, 2007, <https://doi.org/10.1111/j.1365-2753.2006.00671.x> [†]
- [24] C. Tucker et al., “A data mining methodology for predicting early stage Parkinson's disease,” *IIE Transactions on Healthcare Systems Engineering*, vol. 5, no. 4, pp. 238–254, 2015, <https://doi.org/10.1080/19488300.2015.1095256> [†]

- [25] M. Bhat, M. Rabindranath, B. S. Chara, and D. A. Simonetto, "Artificial intelligence, machine learning, and deep learning in liver transplantation," *Journal of Hepatology*, vol. 78, no. 6, pp. 1216–1233, 2023, <https://doi.org/10.1016/j.jhep.2023.01.006> [↑]
- [26] K. M. Pedro, M. A. Alvi, and M. G. Fehlings, "Empirical application of a multidimensional approach to capture clinical benefits," *EBioMedicine*, vol. 108, Art. no. 105365, 2024, <https://doi.org/10.1016/j.ebiom.2024.105365> [↑]
- [27] R. A. Ford, W. Price, and I. I. Nicholson, "Privacy and accountability in black-box medicine," *Michigan Telecommunications and Technology Law Review*, vol. 23, no. 1, 2016. [Online]. Available: <https://repository.law.umich.edu/mttlr/vol23/iss1/1> [↑]
- [28] P. Geurts, A. Irrthum, and L. Wehenkel, "Supervised learning with decision tree-based methods in computational biology," *Molecular Biosystems*, vol. 5, no. 12, pp. 1593–1605, 2009, <https://doi.org/10.1039/B907946G> [↑]
- [29] S. García, J. Luengo, J. A. Sáez, V. López and F. Herrera, "A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 734–750, 2013, <https://doi.org/10.1109/TKDE.2012.35> [↑]
- [30] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling," *Healthcare Analytics*, vol. 3, 100130, 2023, <https://doi.org/10.1016/j.health.2022.100130> [↑]
- [31] R. K. Zimmerman et al., "Proposed clinical indicators for efficient screening for COVID-19 using CART analysis," *Human Vaccines & Immunotherapeutics*, vol. 17, no. 4, pp. 1109–1112, 2021, <https://doi.org/10.1080/21645515.2020.1822135> [↑]
- [32] A. S. Saud, S. Shakya, and B. Neupane, "Analysis of depth of entropy and GINI index based decision trees for predicting diabetes," *Indian Journal of Computer Science*, vol. 6, no. 6, pp. 19–28, 2021, <https://doi.org/10.17010/ijcs/2021/v6/i6/167641> [↑]
- [33] J. Sun, N. Jiang, G. Sun, and W. Huang, "Analysis of CART Algorithms in data mining," in *2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM)*, Jiuzhaigou, China, 2023, pp. 548–553, <https://doi.org/10.1109/MLCCIM60412.2023.00088> [↑]
- [34] M. Bergrab and C. Aßmann, "Automated Bayesian variable selection methods for binary regression models," *AStA Wirtschafts- und Sozialstatistisches Archiv*, vol. 18, no. 2, pp. 203–244, 2024, <https://doi.org/10.1007/s11943-024-00345-1> [↑]