Machine Learning for the Classification of Surgical Patients in Orthodontics

Carlos Andrés Ferro-Sánchez[1] [0009−0004−4872−9869], Christian Orlando Díaz-Laverde[2] [0000−0003−0776−5404], Victor Romero-Cano[1] [0000−0003−2910−5116], Oscar Campo[1] [0000−0002−5007−9613] and Andrés Mauricio González-Vargas[1] [0000−0001−6393−7130]

[1] Universidad Autónoma de Occidente, Faculty of Engineering, Cali, Colombia,
[2] Universidad del Valle, Health Faculty, Cali, Colombia,
[1] Corresponding Author: amgonzalezv@uao.edu.co

Abstract. Dentofacial anomalies, also known as malocclusions, are alterations with a congenital, traumatic, or growth origin. These anomalies can generate functional and aesthetic problems in those who suffer from them and have been reported by the World Health Organization as the third most prevalent oral disease. The most commonly used methods for correcting these anomalies are orthodontics and orthognathic surgery. The diagnosis, and the correct selection of the treatment to be carried out, are part of an extensive process that involves collecting different cephalometric and clinical data, and depend on the clinician's experience. Therefore, no standardized process allows the classification or diagnosis among patients who achieve the best result with orthodontics, that is, non-surgical procedures or if surgical intervention is necessary. This study aims to propose a digital tool based on machine learning algorithms that may help the clinician to select an orthodontics or surgical treatment for patients who are about to start their treatment.

Keywords: Orthodontic, Machine learning, Malocclusion, Surgical, Cephalometric, Classification

1 Introduction

Malocclusion is an anomaly that is characterized by the alteration of craniofacial growth, or the presence of a poor relationship or misalignment between the upper and lower dental arches concerning the transverse or vertical anteroposterior planes [1], which can generate functional problems, aesthetic and psychosocial, and affect social development or emotional wellbeing in both children and adults [2]

Currently, the most widely used methods for correcting malocclusions are orthodontic treatments, sometimes combined with orthognathic surgery [3], depending on the severity of the malocclusion and its classification. Edward Angle proposed a classification of malocclusions based on the anteroposterior relationship of the upper and lower buccal segments [4]:

- Class I: A normal anteroposterior relationship prevails.
- Class II: characterized by mandibular retrognathism and maxillary prognathism. This classification has two subdivisions:
  - Excessive distance (overjet) between the upper and lower incisors.
  - Existence of retroclination of maxillary central incisors.
- Class III: Characterized by prognathism of the mandibular segment.

Since 1989 [5], the World Health Organization (WHO) has reported malocclusions as the third most prevalent oral disease, after caries and periodontal disease [6]; likewise, it affects about 50% of the world population [7]. Recent studies reported the global prevalence of malocclusion in permanent dentition in class I at 74.7%, class II at 19.56%, class III at 5.93%, deep bite at 21.98%, open bite at 3.97% and posterior crossbite at 9.39% [8]. In Latin America, the Pan American Health Organization reported a prevalence and incidence of malocclusions greater than 85% in the population [9].

Diagnosis or problem definition and treatment planning are the most important steps in the correction of malocclusions. Unfortunately, bite correction does not always lead to correction of facial esthetics, and sometimes, facial imbalances occur in the desire to correct the bite [10], [11]. For this reason, if not done correctly, it can end up in extensive treatments that can generate repercussions such as root resorption and increased sensitivity to pain, in addition to affecting esthetics. Therefore, if the diagnostic results indicate that the patient's desired results are not achievable with orthodontic treatment alone, orthognathic surgery or a combination of both should be considered as a therapeutic method [12], as is the case in borderline patients for whom orthodontic treatment is chosen due to cultural, esthetic and financial conditions, even if they are surgical cases[13].

Decision making, or classification of patients according to the required treatment, is an exhaustive process that requires the organization of different diagnostic data, prior knowledge, and experience of the clinician. Therefore, there is no standardized way of doing it. If we could have a tool based on artificial intelligence that helps decision making, it could help the workflow of orthodontists with a high casuistry, as well as help those who still do not have a lot of experience. Currently, some expert algorithms have been developed, capable of reproducing an expert's classification or decision making capacity [12], [14] with an accuracy of 96% [12]. In another study to create a layered system for classifying malocclusions, according to Angle's classification, authors used logistic regression, K-nearest neighbors, random forests, and Bayesian classifiers, and attained accuracies of 88.89%, 83.33%, 88.89%, and 55.66%, respectively [15]. However, no previous study currently allows the classification between surgical and non-surgical patients in the Colombian population.

This study evaluates the feasibility of obtaining the classification of surgical and non-surgical patients with a sample of Colombian patients who are in the process of diagnosis or undergoing orthodontic treatment. In addition, machine learning algorithms focused on binary classification are proposed. First, we present the methodology for acquiring and classifying patients. Second, the process of cleaning and analyzing the data used to train the machine learning algorithms used in the study is presented. Finally, we discuss the results and present some conclusions about the work.

2 Materials and Methods

2.1 A. Data acquisition

The cephalometric data necessary for this study were obtained from 104 cephalometries taken sequentially and chronologically in a dental office in Cali, Colombia. The location of the craniometric and cephalometric points was done manually in the NTN viewer software. These are calculated according to the bony structures of the skull of the patients. This procedure was performed with the help of 2 experienced orthodontists and the process was verified again 4 days later with the same operators until an adequate Cohen's Kappa coefficient was obtained, in this process the patients for whom there were discrepancies between the two specialists were eliminated, therefore the final sample was 86 patients. The cephalometric measurements consisted of 8 angular and 8 linear for a total of 16 measurements that provide sufficient information to determine the dental and skeletal characteristics that are the object of this study (Fig. 1). The measurements used for the characterization of the radiographs are described below.

- Linear
  - Overbite: Vertical overlap of teeth, measured between the upper and lower incisal edges.
  - Overjet: Horizontal overlapping of the teeth, measured from the incisal edge of the upper incisor to the buccal surface of the lower incisor.
  - Spee curve (depth): Occlusal curvature was observed in the sagittal view of the lower arch; for this study, the depth was measured with respect to the occlusal plane.
  - U1-NA: Relationship of the maxillary central incisor with the reference line N-A. Distance from the labial surface of the incisor anterior to the N-A line.
  - L1-NB: Relationship of the mandibular central incisor with the reference line N-B. Distance from the labial surface of the incisor anterior to the N-B line.
  - UL-EP: Distance from the upper lip to the E line traced between the E and pogonion (Pg) points of soft tissues.
  - LL-EP: Distance from the lower lip to the E line drawn between the E points and the soft tissue pogonion.
  - L1-APg: Relationship of the mandibular incisor concerning the line between point A and pogonion.

- Angular:
  - IMPA: Angle formed between the lower incisor and the mandibular plane.

- Upper incisor to palatal plane (UIPP)
- FMIA: Frankfort to lower incisor.
- FMA: Angle formed between the Frankfort plane and the mandibular plane.
- SNA: Angle formed between the saddle-nasion points (N) and point A, which refers to the maxilla's horizontal position with respect to the skull's base.
- SNB: Angle formed between the saddle-nasion points and point B, which refers to the mandible's horizontal position with respect to the skull's base.
- U1-NA: Relationship of the maxillary central incisor with the reference line N-A. The inclination of the axis of the maxillary incisor.
- L1-NB: Relationship of the mandibular central incisor with the reference line N-B. Mandibular incisor axis inclination.

These samples went through an anonymization process where personal data was eliminated. Later, they were classified with the help of an orthodontist with more than ten years of experience, who labeled them as surgical and non-surgical, distributed in 49 samples for surgical and 37 for non-surgical.

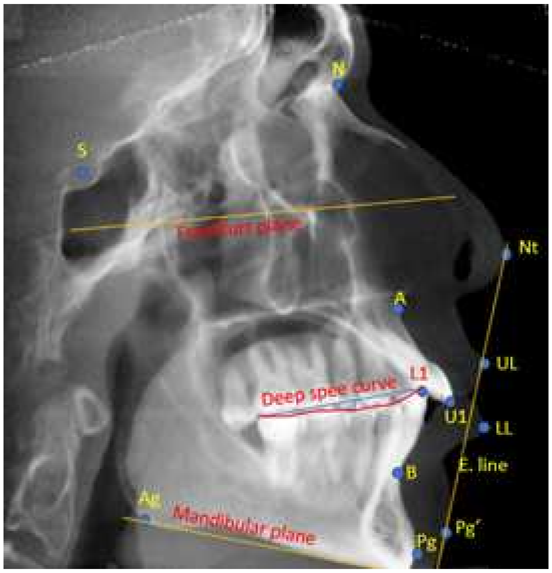

Fig. 1: Location of cephalometric marks.

2.2 Data preprocessing

Once the data set was collected and labeled, it was analyzed. It was observed that there were missing data for some angular and linear measurements. We also found an imbalance in the classes. Because it is a small database, eliminating the samples with missing data is not recommended since this may sacrifice the representativeness of the available data. Therefore, a data imputation process was carried out using the KNNImputter model from the *sklearn* library, which allows imputation to complete the missing values using a K-nearest neighbors-based methodology, avoiding altering the normal distribution of the data. For this step, it was necessary to divide the database into a training set (80%) used to train and validate using a 5-fold cross-validation, and a testing set (20%), using the train_test_split function of the sklearn library to prevent the model from knowing all the data during the test. It was identified that much of the missing data belonged to the surgical class, and there were only two samples of the non-surgical class with missing data in the characteristic overbite. The missing data were located in the characteristics UIPP, Overbite, Overjet, Spee curve, and L1-APG, all with 27 missing data except for overbite, which presented 29 missing data. Once the data samples were imputed, the first training set was generated.

The training dataset distribution presented and imbalance between surgical (n = 38) and non-surgical (n = 30), which could affect the training of machine learning algorithms. To balance the training set, and because it is a small training set, it was decided to carry out an oversampling, using two techniques applied to copies of the original set, from which two training test datasets were obtained:

- Dataset with random oversampling: this dataset was built using random oversampling by resampling, which is based on the random selection of examples of the minority class, to which it makes a smooth replacement and adds them to the training set.
- Synthetic Random Minority Oversampling (SMOTE) dataset: This was obtained using a model that works by selecting a random point from the minority class and calculating the K nearest neighbors for the selected point. The newly added data is selected between the selected point and its neighbors.

Once this process was finished, three training sets were obtained, which were used to train and evaluate the different classifier machine learning algorithms.

2.3 Classification models

In order to perform the classification task, different classifier algorithms were used, such as K-Nearest Neighbors (KNN), Support Vector Classifiers (SVC), Logistic Regression (LR), Decision Trees (DT), Random Forests with and without pre pruning (RF), Bayesian classifiers (NB), gradient boosting for classification (GB), and multilayer perceptron (MLP).

For the selection of features, we used the SelectPercentile and f_classif function of the sklearn library. This method uses a univariate statistical test to select the best features according to the requested percentile. The percentiles 5, 25, 50, 75 and 95, were tested for the different models and trained using the pipeline function and GridSearchCV to obtain the features that allowed the best performance for each independent model.

Because the samples of both classes present a high dispersion, the application of scaling was tested using the StandardScaler function, Robust scaler function and the MaxAbsScaler function. This process was configured within a pipeline in the input data processing stage, where the GridserachCV function selects the scaling that gives the best performance for the evaluated model.

For the construction of the models, a group of hyperparameters was tested (Table 1), using GridserchCV to find the values that would achieve the best performance for each model in accuracy, sensitivity, and f1_score metrics.

Table 1: Models and hyperparameters

Model	Algorithm	Hyperparameters
KNN	KNeighbors Classifier	N_neighbors , weights
SVM	SVC	Kernel, gamma, C
LG	Logistic Regression	C, penalty, max_iter,solver.
RF with prepruning	RandomForest Classifier	N_stimators, max_features, max_depth, criterion
RF	Random Forest Classifier	N_stimators, max_features, criterion
NB	GaussianNB	Var_smoothing
DT with prepruning	Decision Tree Classifier	Criterion, Max_depth
DT	Decision Tree Classifier	Criterion, min_samples-leaf
GB	Gradient Boosting Classifier	N_stimators, max_features, max_depth, criterion
MLP	MLP Classifier	Max_iter, activation, hidden_layer_size, solver

2.4 Model testing

To test the models, the metrics precision, recall, f1 score, and accuracy were used since they are widely common measures to evaluate the performance of binary classifiers [16] in supervised machine learning algorithms. In addition, the metric AUC allows us to evaluate the quality of the classification models and thus choose the best model to use [17].

3 Results

Eight classification models were trained with the three different training datasets, one of which is affected by class imbalance, another balanced using random oversampling, and finally, a balanced dataset using SMOTE oversampling. These datasets provide the best combination of the most important characteristics, the need for scaling or not of the data, and the selection of the best combination of hyperparameters for each model.

The test of these models showed that 75% of them presented better performance with the 16 initial features, which were SVM, LG, RF, DT, NB, and GB 25% used all except the linear measure U1-NA (Table 2).

Several of the algorithms provided acceptable performance (Table 3). However, those with the best performance were the decision trees and gradient boosting for classification, which correctly classified all 10 samples out of 11 belonging to the surgical class in the test set. However, in the case of the non-surgical class, all 7 samples were correctly classified (Table 4).

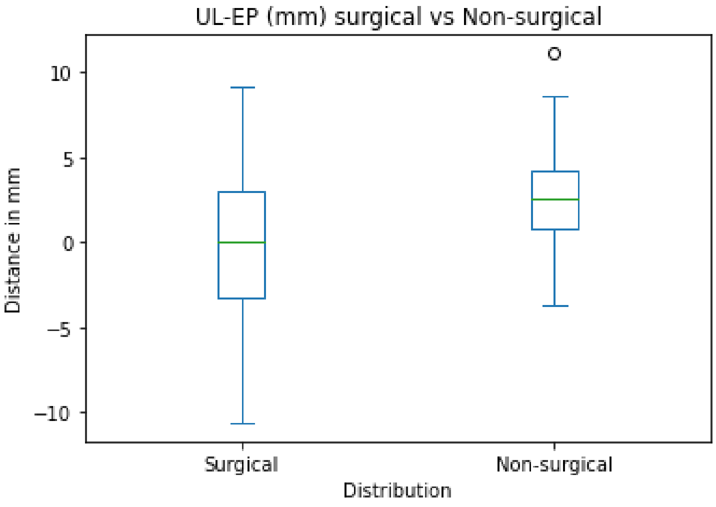

Fig. 2: Distribution of UL-EP measurement among classes

The other algorithms that presented performance of interest were RF, SVM, and MLP, which presented interesting results in most of the metrics, with a greater focus on the AUC metric.

Table 2: Features and hyperparameters selected

Model	Best hyperparameters	Percentile features	Features selected	Scaler	Train set
KNN	n_neighbors = 1, weights = ’uniform’	95	All, except lineal measurement U1-NA	Robust Scaler	Random oversampling
SVM	C=10, gamma=”Scale”	All features	All features	Robust Scaler	Random oversampling
LG	max_iter = 10000, penalty=’l2’, solver = ’lbfgs’, C = 1.0	All features	All features	None	Random oversampling
RF	max_depth=5, max_features = ’None’, n_estimators = 250, criterion: ’gini’	All features	All features	None	Random oversampling
NB	var_smoothing = 0.001232846	All features	All features	MaxAbs Scaler	Random oversampling
DT	max_depth = 50, criterion = ’gini’	All features	All features	None	Random oversampling
GB	criterion = ’friedman_mse’, max_depth = 10, max_features = ’auto’, n_estimators = 10	All features	All features	None	Random oversampling
MLP	hidden_layer_sizes = (75,), max_iter = 3000, activation = ’tanh’, solver = ’adam’	95	All, except lineal measurement U1-NA	None	Random oversampling

Table 3: scores of the metrics of qualification for each model in the test set

Model	Accuracy (%)	Recall (%)	F1_score (%)	Precision (%)	AUC
KNN	66.7	85.7	66.7	54.5	0.72
SVM	88.9	85.7	85.7	85.7	0.97
LG	66.7	85.7	66.7	54.5	0.69
RF	88.9	100	87.5	77.8	0.94
NB	55.6	57.1	50	44.4	0.7
DT	94.4	100	93.3	87.5	0.955
GB	94.4	100	93.3	87.5	0.955
MLP	83.3	100	82.4	0.7	0.92

Table 4: scores for DT qualification metrics

	Precision	Recall	f1_score	Support
Surgical	1	0.91	0.95	11
Non-surgical	0.88	1	0.93	7
Accuracy			0.94	18
Macro avg	0.94	0.95	0.94	18
Weighted avg	0.95	0.94	0.94	18

4 Discussion

Although there is currently a system that handles the classification of orthodontic surgical patients with an accuracy of 96% [12], this study had samples obtained from 316 patients of Korean nationality, and with exclusion criteria for missing teeth, malformed teeth, history of orthodontic treatment, skeletal asymmetries and maxillofacial deformities, which benefits the quality of the data. They also used 12 cephalometric measurements, in addition to six other clinical indices. Our study was performed based on cephalometric measurements taken from Colombian patients, in which there is a wide racial variety [18] which are represented in variations in dental and bone relationships [19], [20], which represented a great variation in the cephalometric measurements [21], [22], despite, some of the algorithms presented good accuracy and sensitivity (Accu=94.4% and Recall = 100%) when classifying the test samples. It was possible to observe the importance of measures that relate to the state of the soft tissues, such as the measurement of the upper lip with Ricketts' E-line (UL-EP), where a certain increase in the projection of the upper lip can be glimpsed in surgical patients, that is, the upper lip extends beyond the margin of the E-line, resulting in negative measurements (Fig. 2). To better characterize both classes, it is necessary to acquire a larger data set that allows a better representation of the problem, and in addition to this, the use of more cephalometric measurements that are indicative of the patient's soft tissue status. In future experiments, we plan to collect more samples, apply more complex algorithms, extract more important data in the diagnostic process and even explore the use of methods based on convolutional neural networks [18], [21].

5 Conclusions

This study evaluated the possibility of generating a system that helps orthodontic clinicians to select a treatment among orthodontics or orthognathic surgery, using cephalometric measurements and a sample of patients of Colombian nationality. The results show good performance of the selected algorithms since they showed an acceptable sensitivity and generalization in the classification, which can be very useful for clinicians in their decision making.

The acquisition of a greater amount and type of data could open the way to use other types of more complex algorithms such as convolutional networks for the extraction of the features embedded in the data.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Z. Zhou et al. “Prevalence of and factors affecting malocclusion in primary dentition among children in Xi’an, China”. en. In: *BMC Oral Health* 16.1 (Sept. 2016). DOI: 10.1186/S12903-016-0285-X..

[2] A.B. Almeida et al. “Dissatisfaction with dentofacial appearance and the normative need for orthodontic treatment: determinant factors”. en. In: *Dental Press J Orthod* 19.3 (May 2014), pp. 120–126. DOI: 10.1590/2176-9451.19.3.120-126.OAR..

[3] Z. Jawad, C. Bates, and T. Hodge. “Who needs orthodontic treatment? Who gets it? And who wants it?” en. In: *British Dental Journal* 218.3 (2015), pp. 99–103. DOI: 10.1038/sj.bdj.2015.51..

[4] J.F. Gravely and D.B. Johnson. “Angle’s classification of malocclusion: an assessment of reliability”. en. In: *Br J Orthod* 1.3 (1974), pp. 79–86. DOI: 10.1179/BJO.1.3.79..

[5] R.R. Santos et al. “Prevalence of malocclusion and related oral habits in 5- to 6-year-old children”. en. In: *Oral Health Prev Dent* 10.4 (2012), pp. 311–8. DOI: 10.3290/J.OHPD.A28901..

[6] N. Cenzato, A. Nobili, and C. Maspero. *Prevalence of Dental Malocclusions in Different Geographical Areas: Scoping Review*. en. Ed. by J. Dent. Oct. 2021. DOI: 10.3390/DJ9100117..

[7] Myriad Edition. *El Desafío de las Enfermedades Bucodentales – Una llamada a la acción global*. es. Accessed: May 21, 2022. [Online]. Available: Ginebra, Apr, 2015. URL: https://www.fdiworlddental.org/sites/default/files/2021-03/book_spreads_oh2_spanish.pdf.

[8] M.S. Alhammadi et al. "Global distribution of malocclusion traits: A systematic review". fr. In: *Dental Press J Orthod* 23.6 (2018), p. 40 1. DOI: 10.1590/2177-6709.23.6.40.E1-10.ONL..

[9] P. Kiep. "Grado de maloclusiones según el índicede estéti-cadental en pacientes que acudieron a la Universidad del Pacífico". es. In: *Rev. cient. cienc. salud* 3.1 (2021), pp. 56–62. DOI: 10.53732/rccsalud/03.01.2021.56..

[10] G.W. Arnett and R.T. Bergman. "Facial keys to orthodontic diagnosis and treatment planning. Part I". en. In: *Am J Orthod Den-tofacial Orthop* 103.4 (1993), pp. 299–312. DOI: 10.1016/0889-5406(93)70010-L..

[11] J.B. Paiva et al. "Facial harmony in orthodontic diagnosis and planning". en. In: *Braz Oral Res* 24.1 (2010), pp. 52–57. DOI: 10.1590/s1806-83242010000100009..

[12] H. Choi. "Artificial Intelligent Model with Neural Net-work Machine Learning for the Diagnosis of Orthognathic Surgery". en. In: *Journal of Craniofacial Surgery* 30.7 (Oct. 2019), pp. 1986–1989. DOI: 10.1097/SCS.0000000000005650..

[13] C. Incorvati et al. "Current Trends in Skeletal Borderline Patients: Surgical versus Orthodontic Treatment Deci-sionsmdash;What Is the Evidence?" en. In: *Applied Sciences* 12.9 (2022), p. 4636. DOI: 10.3390/APP12094636..

[14] K.F. Hung et al. "Potential and impact of artificial intelligence algorithms in dento-maxillofacial radiology". en. In: *Clinical Oral Investigations* (2022), pp. 1–21. DOI: 10.1007/S00784-022-04477-Y..

[15] A.M. Imanthi et al. *Prediction of Malocclusion Pattern of the Orthodontic Patients using a Classification Model*. en. 2019.

[16] D. Chicco and G. Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". en. In: *BMC Genomics* 21.1 (Jan. 2020), pp. 1–13. DOI: 10.1186/S12864-019-6413-7/TABLES/5..

[17] K. Hajian-Tilaki. "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation". en. In: *Caspian J Intern Med* 4.2 (2013). Ac-cessed: May 24, 2022. [Online]. Available: /pmc/articles/PMC3755824/, p. 627.

[18] E. Schwartz-Marín and P. Wade. "Explaining the visible and the invisible: Public knowledge of genetics, ancestry, physical appearance and race in Colombia". en. In: *Soc Stud Sci* 45.6 (Dec. 2015), pp. 886–906. DOI: 10.1177/0306312715621182..

[19] I. Jiménez et al. "Facial growth changes in a Colombian Mestizo population: An 18-year follow-up longitudinal study using linear mixed models". en. In: *American Journal of Orthodontics and Dentofacial Orthopedics* 157.3 (Mar. 2020), pp. 365–376. DOI: 10.1016/J.AJODO.2019.04.032..

[20] L. Aguirre et al. "Frequency and Variability of Five Non-Metric Dental Crown Traits in the Primary and Permanent Dentitions of a Racially Mixed Population from Cali, Colombia". en. In: *Dental Anthropology Journal* 19.2 (Sept. 2006), pp. 39–48. DOI: 10.26575/DAJ.V19I2.119..

[21] Faraj Behbehani et al. "Racial variations in cephalometric analysis between Whites and Kuwaitis". eng. In: *The Angle Orthodontist* 76.3 (May 2006),

pp. 406–411. ISSN: 0003-3219. DOI: 10.1043/0003-3219(2006)076[0406:RVICAB]2.0.CO;2.

[22] J. J. Lee, S. G. Ramirez, and M. J. Will. "Gender and racial variations in cephalometric analysis". eng. In: *Otolaryngology–Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery* 117.4 (Oct. 1997), pp. 326–329. ISSN: 0194-5998. DOI: 10.1016/S0194-5998(97)70121-9.