



Use of automated artificial intelligence to predict the need for orthodontic extractions

Alberto Del Real^{1a,b}
Octavio Del Real^{a,b}
Sebastian Sardina^c
Rodrigo Oyonarte^{a,b}

¹Graduate Orthodontic Program,
Discipline of Orthodontics, Faculty of
Odontology, Universidad de los Andes,
Santiago, Chile

^bPrivate Practice, Santiago, Chile

^cDepartment of Computer Science,
School of Computing Technologies,
RMIT University, Melbourne, Australia

Objective: To develop and explore the usefulness of an artificial intelligence system for the prediction of the need for dental extractions during orthodontic treatments based on gender, model variables, and cephalometric records.

Methods: The gender, model variables, and radiographic records of 214 patients were obtained from an anonymized data bank containing 314 cases treated by two experienced orthodontists. The data were processed using an automated machine learning software (Auto-WEKA) and used to predict the need for extractions. **Results:** By generating and comparing several prediction models, an accuracy of 93.9% was achieved for determining whether extraction is required or not based on the model and radiographic data. When only model variables were used, an accuracy of 87.4% was attained, whereas a 72.7% accuracy was achieved if only cephalometric information was used. **Conclusions:** The use of an automated machine learning system allows the generation of orthodontic extraction prediction models. The accuracy of the optimal extraction prediction models increases with the combination of model and cephalometric data for the analytical process.

[Korean J Orthod 2022;52(2):102-111]

Key words: Extraction vs. non-extraction, Computer algorithm, Decision tree, Orthodontic Index

Received May 11, 2021; Revised September 29, 2021; Accepted October 13, 2021.

Corresponding author: Rodrigo Oyonarte.

Full Professor and Graduate Orthodontic Program Director, Discipline of Orthodontics, Faculty of Odontology, Universidad de los Andes, Mons. Alvaro del Portillo 12.455, Las Condes, Santiago 7620086, Chile.

Tel +56999697644 **e-mail** royonarte@miuandes.cl

How to cite this article: Del Real A, Del Real O, Sardina S, Oyonarte R. Use of automated artificial intelligence to predict the need for orthodontic extractions. Korean J Orthod 2022;52:102-111.

© 2022 The Korean Association of Orthodontists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Malocclusion is a dentofacial anomaly that affects occlusal function, esthetics, and quality of life,¹ and orthodontics is the discipline that encompasses the evaluation, prevention, and treatment of malocclusions. Several methods for identifying the possible causes of malocclusions and treatment approaches have been developed. In recent years, computational approaches, using software systems, have been used in medicine and dentistry to facilitate more efficient diagnostic strategies and better therapeutic results guided by prognostic predictions.² Computational approaches have also been used to quantify the subjective impressions of the expert professional, incorporating the expert's clinical perspective into the systems and making that available to the less experienced doctors.^{3,4} Among these computational techniques, the ones that have received special attention are those from the field of Artificial Intelligence (AI), specifically, machine learning (ML), whereby a computer is *trained* to generate a customized model based on a given dataset, which is used for predictions.² Widely used examples of complex ML algorithms are neural networks and deep neural networks.

As with several advanced computational techniques, applying ML effectively is an intellectually and technically challenging enterprise. ML requires an expert to organize, analyze, and optimize the data and the models and prevent overfitting.⁵ Overfitting is a phenomenon by which excessive iterative learning can increase the goodness-of-fit of the model for the training dataset,⁴ with a decrease in the accuracy of the model when applied to the test set or an external database. It is classically handled by splitting the sample into training and test parts, cross-validation, or more complex but more reliable methods as nested cross-validation, consisting of an inner circle of cross-validation nested within an outer circle of cross-validation. The inner circle is equivalent to the validation set and is used for model selection and optimization, while the outer circle is equivalent to the test set and is used for error estimation. Using nested cross-validation has been accepted as a viable approach to optimally unbiased performance evaluation on the training set while reducing overfitting.^{6,7}

To address the inherent complexity of ML algorithms, Automated ML (AutoML) systems have been developed. AutoML differs from ML systems in that they aim to automatically select, compose, and optimize various ML algorithms for optimal performance for a specified outcome variable. The availability of such systems has facilitated access to AI methods and technology by non-experts.^{5,8,9} Several AutoML systems are available, including IBM AutoWatson (IBM Corp., Armonk, NY, USA) and Auto-WEKA (<https://github.com/automl/autoweika>).⁹

Such systems differ mostly in their optimization techniques or user interfaces.¹⁰ The use of AutoML systems within health care can improve health outcomes, reduce costs, and advance clinical research.¹¹ One of the main benefits of using AutoML systems over conventional systems is their ease of use. This allows non-data science professionals, such as most orthodontic clinicians, to create, test, and run AI systems in their practices without necessarily requiring a highly trained expert to assist them. However, more work needs to be done for the widespread adoption of this technology by healthcare professionals.¹¹

The applications of AI in dentistry are rapidly expanding. AI techniques and methods have been used to identify caries in radiographs¹² and predict periodontal¹³ and endodontic treatment outcomes,¹⁴ among others. In orthodontics, AI systems have been developed for automatic cephalometric tracing, automated diagnosis,¹⁵ growth prediction,¹⁶ treatment outcome prediction,¹⁷ and cervical maturation determination.¹⁸ To date, 7 papers have reported the development of ML prediction models for orthodontic extractions. Six of these papers reported accuracies of 80%,³ 87.5%,¹⁹ 90.5%,²⁰ 93%,⁴ 94.6%,²¹ and 81%,²² respectively. The seventh article reported several accuracies, according to the model and test, ranging from 65 to 98%.²³ The usefulness of AutoML systems has not been assessed for the generation of models for the prediction of the need for orthodontic extractions.

This study aimed to generate prediction models for the need for dental extractions for orthodontic treatment based on gender, model features, and cephalometric records using an AutoML.

MATERIALS AND METHODS

The sex and clinical data of the patients and cephalometric data on comprehensive orthodontic treatments at an orthodontic clinic were obtained from an anonymized data bank of orthodontically treated cases between January 2018 and September 2019 from a single clinical practice run by two orthodontists in Santiago, Chile. These two experienced orthodontic practitioners, with more than 20 and 40 years of exclusive dedication to orthodontics, respectively (worked together for 18 years), performed the diagnosis, treatment planning, and treatment of all the individuals included in this sample. The data were anonymized and processed in a spreadsheet before the conception of this study as follows: once each patient had their orthodontic appliances removed, the orthodontist in charge entered the initial clinical information, including the model and radiographic data and whether or not extractions were performed in the said case into a spreadsheet, avoiding any data that would make the patient identifiable.

Approval for the use of the data from an anonymized databank was obtained from the Research Committee of the Faculty of Odontology at Universidad de los Andes (CPI ODO 26).

The inclusion criteria included receiving consecutive treatments encompassing comprehensive orthodontic treatment in permanent dentition, using either buccal or lingual orthodontic fixed appliances. Patients who had incomplete records or received orthognathic surgical treatment, first-phase treatment, had one or more teeth other than the third molars absent at baseline, or presented with congenital malformations were excluded from the study.

The features used for the development of the optimal predictive models included sex, model variables, and cephalometric data and are shown in Table 1. The data obtained from the anonymized data bank included seven variables considered relevant for predicting the need for extractions. The cephalometric data included in the anonymized database were obtained from cephalometric tracings by an experienced operator and reviewed by one of the two experienced orthodontists using Dolphin® Imaging version 11.95 (Dolphin Imaging, Chatsworth, CA, USA). The only dependent variable included was “Extraction” (NO/YES), as described in Table 1.

Three prediction settings were developed: the first incorporated sex and the model and cephalometric data, while the second and third incorporated only the model and cephalometric data, respectively. Each setting was entered into Auto-WEKA as a .csv (comma-separated values) file to optimize accuracy to identify the optimal model for each setting. The memory limit was set to 2 GB, and the time limits set for Auto-WEKA were 5, 15, 30, and 60 minutes and an overnight limit for which the system ran for at least 18 hours. After these settings were programmed, the “Run” button was clicked and the AutoML system was initiated. Auto-WEKA first performs an automated attribute (variable) selection process, involving 2 algorithms with up to 4 hyperparameters that are optimized to obtain the best variables and prevent redundant and/or irrelevant variables.⁹ Auto-WEKA approaches the selected variables of the sample through 37 learning algorithms, each one being evaluated with its respective hyperparameters, which add up to 160.⁹ Each hyperparameter can have multiple values, and the number of iterations may reach several for simulations that may run for several hours. To evaluate the performance, Auto-WEKA performs a 10-fold nested cross-validation system by default. This consists of a process that divides a sample automatically into 10 parts; 9 of them are used for the training/validation set and the 10th is used for a test set. The training/validation set was then subdivided into 10, with 9 parts used as the training set and the remaining one used as the validation set. Each training/

validation set iterated 10 times, to make every set be once the validation set. Also, the division training/validation vs. test set iterated 10 times, to make every partition to be a test set. This method has been described as a viable way of obtaining an optimally unbiased performance while reducing overfitting.^{6,7}

The model with the best accuracy (number of cases correctly classified by the model related to the final decision of the doctor) for each setting was considered optimal. We also recorded the following metrics for each model: algorithm used, sensitivity, false-positive rate (Type I error or 1-Specificity), precision (positive predictive value), F-score (harmonic mean of the precision and sensitivity), Matthew’s correlation coefficient (correlation between the predicted class and reality), area under the receiver operating characteristic (ROC) curve, area under the curve Precision-Recall curve, and Kappa value.

RESULTS

Out of the 314 orthodontic treatments available in the analyzed database, 214 met the inclusion criteria; 44% of these were received by females. Extractions were performed for 38% of the cases. Further details on the sample are presented in Tables 2 and 3.

Five different extraction prediction models were generated per setting (one for each time limit set). The accuracies obtained for the 5-, 15-, 30-, and 60-minute and overnight models were 80.37%, 86.45%, 80.37%, 80.37%, and 93.93%, respectively. The last and best result was facilitated by feature selection, where the AutoML chose the best variables that predicted the outcomes, which in this case were: maxillary arch discrepancy, mandibular arch discrepancy, molar class-modified, Rickett’s maxillary depth, Rickett’s facial axis, cephalometric molar relationship, and upper incisor protrusion. Based on these 7 variables, the system automatically created, using a multilayer perceptron algorithm, a model that was later optimized and tested following the nested cross-validation method.

For the second setting, the accuracies were 87.38%, 81.78%, 81.78%, 79.91%, and 84.11% for the respective time limits; the best accuracy for this model was achieved for the 5-minute time limit. Using a feature selection algorithm, the following variables were used for a logistic model tree algorithm: maxillary arch discrepancy, mandibular arch discrepancy, and molar class-modified.

Finally, the third model had the following accuracies for the respective time limits: 71.96%, 70.56%, 70.09%, 70.56%, and 70.56%. The 5-minute time limit was associated with the highest accuracy, via a Sequential Minimal Optimization algorithm applied after choosing the best features for predicting the outcome. These fea-

Table 1. Clinical and cephalometric variables included in this study

Variable	Description
Sex	Female or male
Model variable	
Overjet (mm)	The distance between the incisal tips of the upper and lower incisors measured in the horizontal plane
Overbite (mm)	The distance between the incisal tips of the upper and lower incisors measured in the vertical plane
Maxillary arch discrepancy (mm)	Maxillary difference between the space available and the sum of the mesiodistal tooth width from second premolar to second premolar
Mandibular arch discrepancy (mm)	Mandibular difference between the space available and the sum of the mesiodistal tooth width from second premolar to second premolar
Anterior bolton discrepancy (mm)	Measure of the difference between the sum of the mesiodistal width of the front six mandibular teeth and the sum of the mesiodistal width of the front six maxillary teeth multiplied by 0.772
Molar class, modified	Numerical quantification in terms of cusps of difference, that is, with Class I being the value zero, a Class II of 1 cusp being a +1 and a Class III of 1 cusp being -1, and there are intermediate values such as, for example, a quarter cusp (0.25), a half cusp (0.5) and three quarters cusp (0.75)
Canine class, modified	Numerical quantification in terms of cusps of difference, that is, with Class I being the value zero, a Class II of 1 cusp being a +1 and a Class III of 1 cusp being -1, and there are intermediate values such as, for example, a quarter cusp (0.25), a half cusp (0.5) and three quarters cusp (0.75)
Sagittal cephalometric variable	
ANB (°)	The angle formed by the Nasion-Point A plane and Nasion-Point B plane
SNA (°)	The angle formed by the Sella-Nasion plane and the Nasion-Point A plane
SNB (°)	The angle formed by the Sella-Nasion plane and the Nasion-Point B plane
Ricketts' facial convexity (mm)	The distance between Point A and the facial plane
Ricketts' maxillary depth (°)	The angle formed by the Frankfort plane and the plane from Nasion to Point A
Ricketts' facial depth (°)	The angle between the facial plane and Frankfort plane
Articular angle (S-Ar-Go) (°)	The angle formed by the Sella-Articulare plane and the Articulare-Gonion plane
Upper gonial angle (Ar-Go-N) (°)	The angle formed by the Articulare-Gonion plane and the Gonion-Nasion plane
Wits analysis (mm)	Distance between the AO and BO Points, along the occlusal plane
Maxillomandibular difference (mm)	Difference between the Condylion-Anterior Nasal Spine distance and the Condylion-Pogonion distance
Cephalometric overjet (mm)	The distance between the incisal tips of the upper and lower incisors measured along the occlusal plane
Molar relationship (mm)	The distance between the distal surfaces of the lower and upper molars measured along the occlusal plane
Vertical cephalometric variable	
Face height ratio (N-ANS/ANS-Me) (%)	Proportion between the Nasion-Anterior Nasal Spine distance and the Anterior Nasal Spine-Menton distance
Gonial angle (Ar-Go-Me) (°)	The angle formed by the Articulare-Gonion plane and the Gonion-Menton plane
Sum of the angles (°)	Sum of the Articulare-Sella-Nasion angle, the Articular angle and the Gonial angle
Lower gonial angle (N-Go-Me) (°)	The angle formed by the Nasion-Gonion plane and the Gonion-Menton plane
Jarabak Index (%)	The ratio of posterior facial height (Sella-Gonion) to anterior facial height (Nasion-Gnathion)
SN-GoGn (°)	The angle formed by the Sella-Nasion plane and the Gonion-Gnathion plane

Table 1. Continued

Variable	Description
PP-MP (°)	The angle formed by the palatal plane and the mandibular plane
FMA (°)	The angle formed by the Frankfort plane and the mandibular plane
Ricketts' facial axis (°)	The angle formed by the Pt Point-Gnathion plane and the Nasion-Basion plane
Cephalometric overbite (mm)	The distance between the tips of the lower and upper incisors measured perpendicular to the occlusal plane
Dental cephalometric variable	
U1-APo (mm)	The distance from the tip of the upper incisor to the "A-Po" plane
U1-PP (°)	The angle between the long axis of the upper incisor and the palatal plane
L1-APo (mm)	The distance from the tip of the lower incisor to the "A-Po" plane
IMPA (°)	The angle between the long axis of the lower incisor and the mandibular plane
Interincisal angle (°)	The angle formed by the long axes of the central incisors
Soft tissue cephalometric variable	
Labial gap (mm)	The distance between the upper and lower Stomions, measured in the vertical plane
Upper lip to SnPg' (mm)	The distance between the upper Labrale and the Subnasal-Soft Tissue Pogonion plane, along the horizontal plane
Lower lip to SnPg' (mm)	The distance between the lower Labrale and the Subnasal-Soft Tissue Pogonion plane, along the horizontal plane
Upper lip to subnasal vertical (mm)	The distance between the upper Labrale and a vertical plane projected over Subnasal Point
Lower lip to subnasal vertical (mm)	The distance between the lower Labrale and a vertical plane projected over Subnasal Point
Upper incisor exposure (mm)	The distance between the tips of the upper incisor and the upper Stomion, along the vertical plane
Outcome variable	
Extractions NO/YES	Extractions done or not in the patient due to orthodontic reasons

tures included Rickett's maxillary depth, Rickett's facial axis, cephalometric molar relationship, upper incisor protrusion, and wits appraisal.

Table 4 shows the other outcomes for each of these settings.

DISCUSSION

This study is the first to explore the performance of an AutoML system for the generation of predictive models for dental extractions for orthodontic treatments using sample sizes comparable to those used for traditional ML systems in the past.^{3,4,19-22} Our inclusion and exclusion criteria were similar to those used in previous studies on traditional ML methods,^{3,4,19-22} which purposely excluded orthognathic patients and treatments for the absence of at least one tooth at baseline. The AutoML technology allowed us to achieve accuracies comparable to those reported in the literature using traditional ML techniques, albeit with a simplified methodology requir-

ing minimal ML expertise.

This study does not advocate for automated decision-making for the extraction or non-extraction of teeth. We consider these AI tools as additional resources for the practitioner and can even be used after an initial judgment, with considerations of some characteristics inherent to these systems. First, ML approaches are relatively black boxes, and the contribution of each feature may be difficult to assess;²⁴ in our case, we determined the variables used, but not the contribution of each of them, despite the significant research advances on making ML models more interpretable.²⁵ The overall decision will take into account other factors that the tool may not have access to, such as age, previous diseases and dental interventions, presence of root resorption, functional factors, growth potential, patient preferences, as well as clinician-related diagnostic and therapeutic orientations, among others. Hence, like in any intervention involving the wellbeing of humans, the doctor's critical judgment should remain at the center of the process. An

Table 2. Detailed description of the sample

Variable	Without extraction	With extraction	p-value	Global
Model variable				
Overjet (mm)	3.17 ± 1.42	3.91 ± 2.25	0.004*	3.448 ± 1.804
Overbite (mm)	3.55 ± 1.83	3.08 ± 1.82	0.069	3.376 ± 1.836
Maxillary arch discrepancy (mm)	0.24 ± 2.86	-2.30 ± 4.03	0.000*	-0.722 ± 3.565
Mandibular arch discrepancy (mm)	-0.09 ± 2.50	-2.17 ± 2.75	0.000*	-0.882 ± 2.786
Anterior bolton discrepancy (mm)	-0.58 ± 1.14	-0.36 ± 1.23	0.185	-0.498 ± 1.170
Molar class, modified	0.22 ± 0.35	0.51 ± 0.39	0.000*	0.329 ± 0.390
Canine class, modified	0.36 ± 0.32	0.52 ± 0.37	0.001*	0.418 ± 0.347
Sagittal cephalometric variable				
ANB (°)	2.62 ± 1.89	3.52 ± 2.39	0.003*	2.962 ± 2.134
SNA (°)	80.87 ± 3.67	80.12 ± 3.64	0.147	80.585 ± 3.671
SNB (°)	78.25 ± 3.41	76.60 ± 3.33	0.000*	77.624 ± 3.466
Rickett's facial convexity (mm)	1.79 ± 2.55	3.09 ± 3.06	0.001*	2.278 ± 2.818
Rickett's maxillary depth (°)	89.05 ± 2.72	87.54 ± 2.44	0.000*	90.562 ± 3.083
Rickett's facial depth (°)	90.71 ± 2.93	90.32 ± 3.32	0.370	88.479 ± 2.710
Articular angle (S-Ar-Go) (°)	146.08 ± 6.71	145.32 ± 6.43	0.415	145.791 ± 6.603
Upper gonial angle (Ar-Go-N) (°)	50.07 ± 4.29	50.03 ± 3.78	0.945	50.057 ± 4.092
Wits analysis (mm)	-0.74 ± 2.78	0.89 ± 3.29	0.000*	-0.125 ± 3.079
Maxillomandibular difference (mm)	32.95 ± 4.83	31.49 ± 4.59	0.030*	32.393 ± 4.748
Cephalometric overjet (mm)	4.23 ± 1.61	4.67 ± 2.22	0.095	4.395 ± 1.873
Molar relationship (mm)	-0.48 ± 1.40	0.61 ± 1.86	0.000*	-0.067 ± 1.668
Vertical cephalometric variable				
Face height ratio (N-ANS/ANS-Me) (%)	82.41 ± 7.04	81.02 ± 6.55	0.152	81.884 ± 6.877
Gonial angle (Ar-Go-Me) (°)	121.75 ± 6.53	123.18 ± 5.16	0.095	122.29 ± 6.077
Lower gonial angle (N-Go-Me) (°)	71.66 ± 4.77	73.16 ± 4.39	0.023*	72.228 ± 4.674
Jarabak Index (%)	65.87 ± 4.85	64.93 ± 4.33	0.154	65.511 ± 4.669
SN-GoGn (°)	28.87 ± 5.86	30.73 ± 5.58	0.023*	29.572 ± 5.814
PP-MP (°)	24.61 ± 5.38	25.99 ± 4.72	0.058	25.132 ± 5.171
FMA (°)	22.51 ± 5.10	24.00 ± 4.50	0.032*	23.073 ± 4.925
Rickett's facial axis (°)	89.17 ± 4.28	87.42 ± 3.59	0.002*	88.509 ± 4.115
Cephalometric overbite (mm)	4.71 ± 4.81	3.55 ± 4.66	0.085	4.267 ± 4.774
Dental cephalometric variable				
U1-APo (mm)	6.18 ± 2.36	7.53 ± 2.83	0.000*	6.693 ± 2.623
U1-PP (°)	109.97 ± 6.14	110.62 ± 7.08	0.479	110.214 ± 6.503
L1-APo (mm)	2.18 ± 2.24	2.94 ± 2.37	0.019*	2.466 ± 2.316
IMPA (°)	93.67 ± 6.17	96.57 ± 7.44	0.002*	94.766 ± 6.812
Interincisal angle (°)	131.76 ± 9.82	126.82 ± 11.46	0.001*	129.891 ± 10.716
Soft tissue cephalometric variable				
Labial gap (mm)	2.25 ± 2.12	2.54 ± 2.63	0.377	2.357 ± 2.322
Upper lip to SnPg' (mm)	4.06 ± 2.09	4.69 ± 2.12	0.035*	4.297 ± 2.118
Lower lip to SnPg' (mm)	3.28 ± 2.32	3.58 ± 2.35	0.362	3.393 ± 2.330
Upper lip to subnasal vertical (mm)	1.94 ± 2.25	2.04 ± 2.68	0.769	1.979 ± 2.416
Lower lip to subnasal vertical (mm)	-2.6 ± 3.34	-3.77 ± 5.36	0.050*	-2.882 ± 3.547
Upper incisor exposure (mm)	4.03 ± 1.99	4.17 ± 2.06	0.623	4.081 ± 2.013

Values are presented as mean ± standard deviation.

*Significance at a 0.05 level.

See Table 1 for definition of each clinical and cephalometric variable.

Table 3. Summarized description of the sample

Variable	Without extraction	With extraction	Total
Sex			
Female	72 (54)	48 (59)	120 (56)
Male	61 (46)	33 (41)	94 (44)
Skeletal sagittal class			
Class I ($0 \leq \text{ANB} \leq 4$)	95 (71)	44 (54)	139 (65)
Class II ($\text{ANB} > 4$)	30 (23)	32 (40)	62 (29)
Class III ($\text{ANB} < 0$)	8 (6)	5 (6)	13 (6)
Skeletal vertical pattern			
Normodivergent ($28 \leq \text{SN-GoGn} \leq 32$)	56 (42)	40 (49)	96 (45)
Hypodivergent ($\text{SN-GoGn} < 28$)	61 (46)	28 (35)	89 (42)
Hyperdivergent ($\text{SN-GoGn} > 36$)	16 (12)	13 (16)	29 (13)

Values are presented as number (%).

See Table 1 for definition of each cephalometric variable.

AI-based system should serve as a complementary piece of information.

For our sample, the prevalence of cases of extractions was 38%, which is expected for an orthodontic clinic and is consistent with those reported in the literature, ranging from 50% in England for patients between 18 and 24 years of age²⁶ to 45.8% in Brazil,²⁷ 40% in China,³ and 25% in the United States.²⁸ Although the non-extraction protocol was dominantly used for our sample, a 40:60 proportion (similar to the 38:62 of our study) is usually accepted for a balanced sample. The imbalance in our sample may be considered minimal.²⁹ The final composition of the sample was determined by the inclusion and exclusion criteria applied to the anonymized database. It was considered that increasing the extraction group could generate a biased sample. Considering the similarities between our sample and that of other studies previously cited on dental extractions and ML,^{3,4} we considered this sample suitable for the development of optimal prediction models.

The prediction of the need for extractions as a YES/NO dichotomous variable based on the combination of the model and cephalometric data showed an accuracy of 93.9% and an F1-value of 0.939 using a multilayer perceptron. This result is consistent with those published by other authors using similar methods but with conventional ML instead of Auto-ML. The reported accuracies ranged from 80 to 94.6%.^{3,4,19-22} The study by Li et al.²¹ reported a sensitivity of 94.6%, specificity of 93.8%, and an area under the ROC curve of 0.982, while the current study achieved 93.9%, 92%, and 0.915, respectively.

Considering the other two settings, which only used model or cephalometric data, the performance was 15.4% better when using only model information than when using only cephalometric information. This sug-

gests that model variables may be more relevant for prescribing extractions than strictly cephalometric data. This may be explained by the fact that tooth extractions are highly affected by maxillary and mandibular arch discrepancies, variables that were included during the feature selection process for both settings 1 and 2. Nonetheless, the best performance was obtained when using both model and cephalometric data, with an accuracy of 6.5% greater than that achieved with the use of only model information and 21.9% greater than that with the use of only cephalometric data. All the variables automatically selected for setting 2 were also included for setting 1. In the same way, all variables chosen for setting 3, excluding one (Wits Appraisal), were also selected by the AutoML system for setting 1, showing high consistency in this regard.

Despite being included in this study, gender was never considered relevant during the feature (variable) selection process by Auto-WEKA. Some variables, such as overjet (OJ) and overbite (OB), were included as both model and cephalometric variables, creating an overlap between these two versions that could traditionally lead to overexpression of this feature in the models. Nevertheless, since the system automatically selects the best features and omits the redundant and/or irrelevant variables, the study considered feeding the Auto-AI system with all the variables available in the database and allow the system to automatically determine which ones to include in the models. For OJ and OB, no final algorithm considered the model or cephalometric data for these variables.

There is no minimum requirement for the sample size required to perform ML and this study used a relatively large database in the context of the published literature in the topic, it is accepted that algorithms reach their

Table 4. Results obtained by Auto-WEKA for the 3 settings studied

Setting	Time limit	Algorithm	Accuracy	Sensitivity	FP-rate	Precision	F-score	MCC	ROCAUC	PR AUC	Kappa
Setting 1 (clinical and Rx data)	5 minutes	Bagging	80.3738	0.804	0.25	0.802	0.8	0.575	0.861	0.859	0.57
	15 minutes	Random Committee	86.4486	0.864	0.174	0.864	0.863	0.709	0.903	0.896	0.706
	30 minutes	Multilayer Perceptron	80.3738	0.804	0.34	0.802	0.802	0.577	0.864	0.863	0.575
	60 minutes	Multilayer Perceptron	80.3738	0.804	0.34	0.802	0.802	0.577	0.864	0.863	0.575
Setting 2 (only clinical data)	Overnight	Multilayer Perceptron	93.9252	0.939	0.08	0.94	0.939	0.87	0.915	0.913	0.869
	5 minutes	LMT	87.3832	0.874	0.173	0.876	0.871	0.73	0.908	0.918	0.723
	15 minutes	REP Tree	81.7757	0.818	0.222	0.816	0.816	0.608	0.822	0.79	0.606
	30 minutes	REP Tree	81.7757	0.818	0.222	0.816	0.816	0.608	0.822	0.79	0.606
Setting 3 (only Rx data)	60 minutes	J48	79.9065	0.799	0.262	0.798	0.794	0.564	0.786	0.754	0.557
	Overnight	Random Tree	84.1121	0.841	0.222	0.845	0.836	0.659	0.898	0.891	0.647
	5 minutes	SMO	71.9626	0.72	0.378	0.715	0.705	0.379	0.735	0.751	0.364
	15 minutes	Multilayer Perceptron	70.5607	0.706	0.392	0.699	0.706	0.691	0.346	0.737	0.755
	30 minutes	SMO	70.0935	0.701	0.39	0.693	0.689	0.338	0.741	0.756	0.329
	60 minutes	AdaBoost	70.5607	0.706	0.368	0.699	0.698	0.355	0.717	0.699	0.351
	Overnight	Bagging	70.5607	0.706	0.406	0.701	0.686	0.343	0.741	0.734	0.324

Accuracy = $(TP + TN) / (TP + TN + FP + FN)$; Sensitivity = $TP / (TP + FN)$; FP-rate = $FP / (FP + TP)$; Precision = $TP / (TP + FP)$;

$$F - score\beta = (1 + \beta^2) \times \frac{Precision \times Sensitivity}{\beta^2 + Precision + Sensitivity}; MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN FN)}}$$

Kappa is the measure of how closely the instances were correctly classified by the algorithm, comparing its accuracy with that of a random classifier. TP, true positive; TN, true negative; FP, false positive; FN, false negative; MCC, Matthews correlation coefficient; AUC-ROC, area under the receiver operating characteristic curve; AUC-PR, area under the Precision-Recall (sensitivity) curve; Rx, radiographic; LMT, logistic model tree; REP, reduced error pruning; SMO, sequential minimal optimization.

greatest potential with greater data sets. From that perspective, the sample for this study was small, as with all other studies published to date in the field. Our sample was obtained from an anonymized databank, and it was not possible to perform a reliability analysis, which is a common issue in the field^{3,4,19-22} and is more evident in our study as two practitioners generated the data. Nevertheless, both doctors had worked together for almost two decades and systematically discussed the diagnosis and treatment plans of complex cases (i.e. any borderline extraction or compensatory orthodontic case). We believe that this favors the consistency of the treatments plans of both doctors. The criterion for this process is based on cephalometric skeletal, dental, and soft tissue readings, as well as clinical aspects that lie within a normal range. Once a diagnosis is reached, treatment planning is performed to modify abnormal occlusal characteristics without negatively affecting the soft tissues of the patient, hence, defining the need for extraction or non-extraction treatment protocols. Both clinicians plan their treatments aiming at optimal occlusal and facial results, using mainly active self-ligating brackets and contemporary orthodontic techniques, including temporary anchorage devices whenever deemed necessary. Therefore, the decision to extract or not to extract depended on the clinical presentation of the dentition combined with skeletal, dental, and soft tissue variables. Whenever possible, the non-extraction approach was preferred, as long as satisfactory occlusal and facial balance could be achieved.

The model generated does not represent the analysis of a particular doctor, but it is a third and singular algorithm, which could not represent the opinion of anyone despite being based on data provided by two orthodontists. In this line, future research should include treatments by several orthodontists, which would lead to the generation of a model with outcomes representing the unbiased recommendations of the professionals included.

Another potential limitation of this study is the degree of overfitting that may have occurred in the models. While the nested cross-validation system greatly reduces overfitting, there can always be a remnant, as with any ML design. This overfitting could explain why some models trained for a few minutes had better performance than those developed for several hours.

Within the decision-making process in Orthodontics, the prescription of extractions is of special relevance, given its irreversible nature as well as the fact that it severely conditions both the occlusal and soft tissue response to treatment. In addition, the indication of extractions is based on several factors, including the clinical, cephalometric, and socio-cultural.³⁰ This multifactorial nature can make clinical decision-making par-

ticularly difficult.

The results of our work suggest a great potential for this readily available AutoML system in the field of orthodontics. Through the development of highly reliable predictive models²¹ (necessarily including all the other variables previously named), AutoML systems can, through a simple methodology and minimal ML expertise, assist clinicians in challenging clinical decision-making scenarios such as tooth extractions.

CONCLUSION

Three different models for the prediction of the orthodontic need of dental extractions were generated and tested using an AutoML method, and they achieved accuracies of up to 93.9% for predicting the need for tooth extractions, similar to those obtained by more complex methods.

Prediction models for the need for dental extractions achieve their best performance when model and cephalometric data are combined, although model data seem more relevant.

The use of AutoML systems simplifies the process of model generation, making it less operator-dependent and allowing the generation of several models for accurate predictions.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Hassibi K. Machine learning vs. traditional statistics: different philosophies, different approaches [Internet]. Issaquah: Data Science Central; 2016 Oct 10 [cited 2019 Oct 25]. Available from: <https://www.datasciencecentral.com/profiles/blogs/machine-learning-vs-traditional-statistics-different-philosophy-1>.
2. Bahaa K, Noor G, Yousif Y. The artificial intelligence approach for diagnosis, treatment and modelling in orthodontic. In: Naretto S, ed. Principles in contemporary orthodontics. London: IntechOpen; 2011. p. 451-92.
3. Xie X, Wang L, Wang A. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. *Angle Orthod* 2010;80:262-6.
4. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop* 2016;149:127-33.

5. Cui C, Wang S, Zhou J, Dong A, Xie F, Li H, et al. Machine learning analysis of image data based on detailed MR image reports for nasopharyngeal carcinoma prognosis. *Biomed Res Int* 2020;2020:8068913.
6. Behr M, Noseworthy M, Kumbhare D. Feasibility of a support vector machine classifier for myofascial pain syndrome: diagnostic case-control study. *J Ultrasound Med* 2019;38:2119-32.
7. Peeken JC, Spraker MB, Knebel C, Dapper H, Pfeiffer D, Devecka M, et al. Tumor grading of soft tissue sarcomas using MRI-based radiomics. *EBioMedicine* 2019;48:332-40.
8. Dinsmore T. Automated machine learning: a short history [Internet]. Boston: DataRobot; 2016 Mar 30 [cited 2019 Nov 13]. Available from: <https://www.datarobot.com/blog/automated-machine-learning-short-history/>.
9. Kotthof L, Thornton C, Hoos HH, Hutter F, Leyton-Brown K. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. *J Mach Learn Res* 2017;18:1-5.
10. Orlenko A, Kofink D, Lyytikäinen LP, Nikus K, Mishra P, Kuukasjärvi P, et al. Model selection for metabolomics: predicting diagnosis of coronary artery disease using automated machine learning. *Bioinformatics* 2020;36:1772-8.
11. Waring J, Lindvall C, Umeton R. Automated machine learning: review of the state-of-the-art and opportunities for healthcare. *Artif Intell Med* 2020;104:101822.
12. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018;77:106-11.
13. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci* 2018;48:114-23.
14. Suebnukarn S, Rungcharoenporn N, Sangsuratham S. A Bayesian decision support model for assessment of endodontic treatment outcome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2008;106:e48-58.
15. Murata S, Lee C, Tanikawa C, Date S. Towards a fully automated diagnostic system for orthodontic treatment in dentistry. Paper presented at: 2017 IEEE 13th International Conference on e-Science (e-Science); 2017 Oct 24-27; Auckland, New Zealand. Piscataway: IEEE, 2017. p. 1-8.
16. Scala A, Auconi P, Scazzocchio M, Caldarelli G, McNamara JA, Franchi L. Complex networks for data-driven medicine: the case of Class III dentoskeletal disharmony. *New J Phys* 2014;16:115017.
17. Auconi P, Scazzocchio M, Cozza P, McNamara JA Jr, Franchi L. Prediction of Class III treatment outcomes through orthodontic data mining. *Eur J Orthod* 2015;37:257-67.
18. Sokic E, Tiro A, Sokic-Begovic E, Nakas E. Semi-automatic assessment of cervical vertebral maturation stages using cephalograph images and centroid-based clustering. *Acta Stomatol Croat* 2012;46:280-90.
19. Martina R, Teti R, D'Addona D, Iodice G. Neural network based system for decision making support in orthodontic extractions. In: Pham DT, Eldukhri EE, Soroka AJ, eds. *Intelligent production machines and systems*. Elsevier Science; 2006. p. 235-40.
20. Takada K, Yagi M, Horiguchi E. Computational formulation of orthodontic tooth-extraction decisions. Part I: to extract or not to extract. *Angle Orthod* 2009;79:885-91.
21. Li P, Kong D, Tang T, Su D, Yang P, Wang H, et al. Orthodontic treatment planning based on artificial neural networks. *Sci Rep* 2019;9:2037.
22. Zaytoun ML. An empirical approach to the extraction versus non-extraction decision in orthodontics [Master's thesis]. Chapel Hill: University of North Carolina at Chapel Hill; 2019.
23. Suhail Y, Upadhyay M, Chhibber A, Kshitiz. Machine learning for the diagnosis of orthodontic extractions: a computational analysis using ensemble learning. *Bioengineering (Basel)* 2020;7:55.
24. Paton C, Kobayashi S. An open science approach to artificial intelligence in healthcare. *Yearb Med Inform* 2019;28:47-51.
25. The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir Med* 2018;6:801.
26. Mew J, Trenouth M. How many teeth are extracted as part of orthodontic treatment? A survey of 2038 UK residents. *Int J Dent Oral Sci* 2018;S1:02:001:1-5.
27. Dardengo Cde S, Fernandes LQ, Capelli Júnior J. Frequency of orthodontic extraction. *Dental Press J Orthod* 2016;21:54-9.
28. Jackson TH, Guez C, Lin FC, Proffit WR, Ko CC. Extraction frequencies at a university orthodontic clinic in the 21st century: demographic and diagnostic factors affecting the likelihood of extraction. *Am J Orthod Dentofacial Orthop* 2017;151:456-62.
29. Imbalanced data [Internet]. Google Developers; [cited 2021 Jul 27]. Available from: <https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data>.
30. Proffit WR, Fields HW, Sarver DM. *Contemporary orthodontics*. 5th ed. St. Louis: Elsevier; 2013.