A Machine Learning Approach to Evaluating the Relationship between Dental Extraction and Craniofacial Growth in Adolescents

Guillermo Hernández[1,2][0000−0002−7481−5961], Alfonso González-Briones[1,2][0000−0002−3444−4393], José Machado[3][0000−0003−4121−6169], Pablo Chamoso[1,2][0000−0001−5109−3583], and Paulo Novais[3][0000−0002−3549−0754]

[1] BISITE Research Group, University of Salamanca. Edificio Multiusos I+D+i, 37007, Salamanca, Spain
{guillehg, alfonsogb, chamoso}@usal.es
[2] Air Institute, IoT Digital Innovation Hub, Carbajosa de la Sagrada, 37188. Salamanca, Spain
[3] Algoritmi Center/LASI, University of Minho, 4710-057 Braga, Portugal
{jmac, pjon}@di.uminho.pt

Abstract. There may be multiple reasons for tooth extraction, such as deep cavities, an infection that has destroyed an important portion of the tooth or the bone that surrounds it, or for orthodontic reasons, such as the lack of space for all the teeth in the mouth. In the case of orthodontics, however, there is a relationship between tooth extraction and the craniofacial morphological pattern. The purpose of this study is to establish whether such a relationship exists in adolescents and to evaluate it and to serve as a tool to support medical decision making.. Machine Learning techniques can now be applied to datasets to discover relationships between different variables. Thus, this study involves the application of a series of Machine Learning techniques to a dataset containing information on orthodontic tooth extraction in adolescents. It has been discovered that by following simple rules it is possible to identify the need of treatment in 98.7% of the cases, while the remaining can be regarded as "limited cases", in which an expert's opinion is necessary.

Keywords: Machine Learning · Craniofacial Morphological Growth · Dental Extraction · Orthodontic treatments.

1 Introduction

To successfully complete orthodontic treatment, tooth extraction may sometimes be necessary. The main reasons for this procedure are: to solve negative osteodental discrepancy (severe dental crowding), camouflage a horizontal skeletal discrepancy (class III or class II) or vertical discrepancy (vertical growth pattern or skeletal open bite), to improve the facial aesthetics of the patient in cases of protrusion or lip incompetence, etc. [7]. The present study determines whether tooth extraction is necessary by examining the craniofacial morphology and the growth pattern. Orthodontists consider different factors when identifying negative osteodental discrepancy. Thus, the rules identified in this paper could support orthodontists' in making decisions. The following factors have been identified:

- Tooth factors: the size of the teeth (macrodontia), number of teeth (supernumerary), pathology (tooth decay, periodontal disease), abnormal position (ectopias, the inclination of the incisors).
- Arcade factors: narrow arcade, osteodental severe discrepancy, changes in the midline.
- Occlusal factors: open bite, molar, and canine class, increased overjet.
- Facial factors: aesthetics protrusive facial profile and vertical growth patterns.
- General factors: patient low growth potential and low degree of patient cooperation.

The presence of crowded and irregular anterior teeth is the most cited reason for seeking orthodontic treatment [25]. Crowding has been associated with vertical growth, lower incisor eruption, and increased vertical dentoalveolar eruption. One might expect crowding and facial divergence to be associated because divergence increases anterior vertical dentoalveolar eruption. Hyperdivergence results in the retroclination of the incisors, which may cause crowding by reducing the arch length.

The following techniques have been employed in this research for data analysis: statistical techniques such as mutual information and correlation analysis; visualization techniques such as kernel density estimation; non-supervised techniques such as association rules; and supervised learning techniques such as decision trees. A series of other techniques have been evaluated, however, it has been decided not to use them due to poor performance or complexity that did not manage to overcome the results obtained.

The remainder of the paper is structured as follows: the next section presents a brief review of related studies, as well as the craniofacial growth patterns that have been considered in this research and the corresponding data set. Section 3 presents some of the widely used techniques to analyse data in the dental sector. Section 4 describes the proposed analysis process. Section 5 presents the results and the conclusions drawn from the conducted research.

2 Biological Background

The experiments have been performed on a set of 303 patients. Nine craniofacial growth patterns were calculated for those patients. This is a group of patients with severe malocclusions presenting a negative osteodental discrepancy of more than 6 mm in the lower arch. These patients were undergoing orthodontic treatment in two private clinics.

The identifiers and patterns used to conduct this study will now be presented. The variables used are shown in parentheses.

- Tooth factors: the size of the teeth (macrodontia), number of teeth (supernumerary), pathology (tooth decay, periodontal disease), abnormal position (ectopias, the inclination of the incisors).
- Age (age): nominal variable of patients between 12 and 15 years of age. The treatment begins at this age.
- Sex (sex): numeric variable where 0 corresponds to a male and 1 to a female.
- Jaw Plane (JaPl) (Tweed, 1946) [27]: numeric variable, in degrees.
  The mandibular plane angle: formed by the mandibular plane and the Frankfort horizontal plane. The Frankfort plane: the point is drawn from the porion point to the orbital point. It is the basic horizontal reference line on the cephalometric tracing.
  Mandibular plane: a tangent to the lower edge of the mandibular connecting the Me point with the lowest point of the mandibular branch. It is a reference that describes the morphology and/or mandibular position. The junction of the front end with the facial plane forms the gnathion cephalometric point.
  Standard value: 26° at 9 years. Decreases 0.3° per year. S. D.: $\pm 4^{\circ}$.
  Interpretation: Low values correspond to brachyfacial patients, with a squared jaw. A high value corresponds to a dolichofacial biotype, and the jaw morphology indicates a very obtuse angle, a clockwise mandibular implant, or a combination of both.
- Vert Ricketts (verRic) (Ricketts, 1956) [21]: numeric variable, in S.D. Facial biotype is determined by the lateral cephalometric VERT index. The type and degree of the vertical growth of the lower third of the face are established numerically by this index. The growth is caused by the anterior or posterior rotation of the mandible. Other parameters such as gender and age are also considered. To obtain the result, five angles are taken into account to position the jaw: face depth, jaw plane, facial axis, lower face height, and jaw arch.
  To Ricketts, VERT is the vertical variation coefficient. It is obtained by calculating the arithmetic mean of the difference between the patient's value and the normal value for that age, divided by the standard deviation. Dolicho standard deviations are characterised by a negative sign. Deviations regarding brachy take a positive sign. If they remain within the norm, the value is 0. To calculate the facial biotype, it is first necessary to add the results of the 5 cephalogram measures. Finally, the resulting number is divided by 5,
  - Low to -1.5: Severe dolicho
  - From -1.5 to -1: Moderate dolicho
  - From -1 to -0.5: Soft dolicho
  - From -0.5 to 0.5: Mesofacial
  - From 0.5 to 1: Soft brachyfacial
  - From 1 to 1.5: Moderate brachyfacial
  - Over 1.5: Severe brachyfacial
- Jarabak Spheres (JaSp): numeric variable in % [16]. (Figure 1). Jarabak identifies three types of growth according to the general growth direction
  A) Counter-clockwise.
  B) Clockwise.
  C) Direct down.

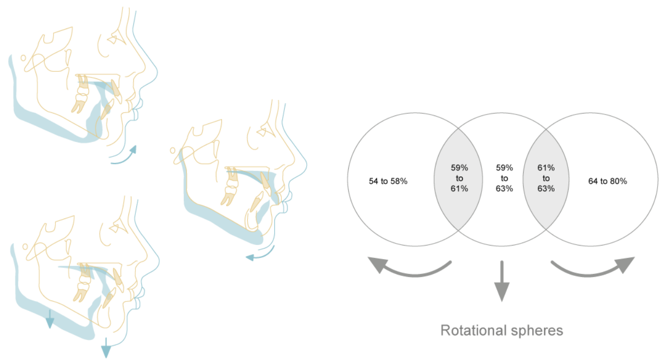


Fig. 1: Rotational Spheres (from Jarabak and Fizzell, 1971)

Patients placed in areas A and C have different growth characteristics, as illustrated in the diagram in Figure 1. In the case of patients in A and C, there is no doubt as to whether the growth pattern is dolichofacial or brachyfacial. However, the growth percentage of the patients in B corresponds to what Jarabak called vertical growth; this type of growth may develop either dolichofacial or barchyfacial pattern, which is determined by the muscles (Figure 2).

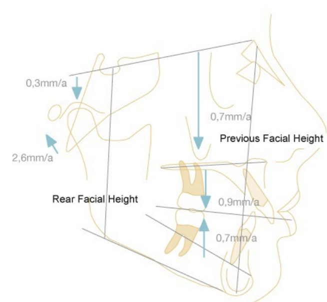
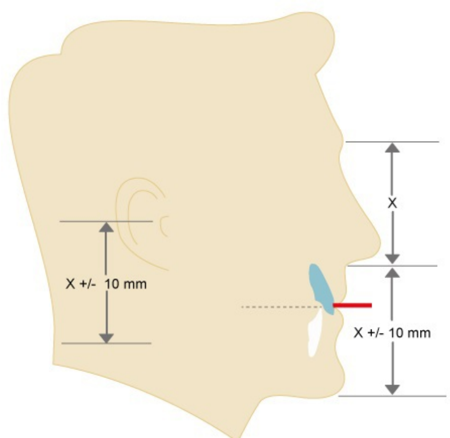


Fig. 2: Rear / Previous Facial Height (from Jarabak and Fizzel, 1971)

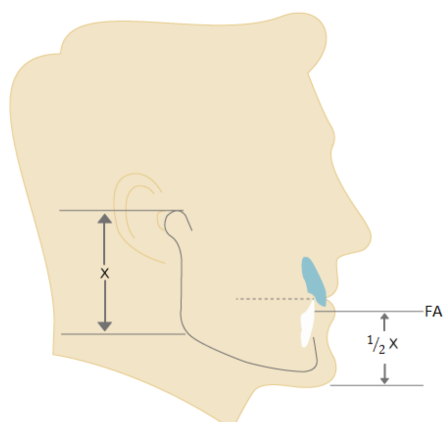
– Previous Facial Height (PrFH) / Rear Facial Height (ReFH): numeric variable in mm [3]. The anterior and posterior numerical facial height values are linear, not angular, and should be proportional to define a balanced face shape. Several reference points are used to measure them: Glabella, subnasal, soft chin, and condylar gonion. Their thirds should be symmetrical as shown in Figure 3 A).

The FA Point is used as a reference for hard tissue. The FA point is also called lower incisor Crown. The distance between this point and the hard tissue chin should be half the lower posterior third. If the face is balanced, the length of the ramus should be twice the height of the symphysis measured from FA to the hard chin (Figure 3 B). If the length of the branch is greater, facial growth will be horizontal, otherwise, it will be vertical.

- Facial Pattern (FacPat): nominal variables can take the following values, 0 for brachy, 1 for meso and 2 for dolicho.
- Type of Treatment (class): Nominal variable where 0 indicates no extraction of tooth and 1 indicates that there has been an extraction.



(a)



(b)

Fig. 3: Anterior /posterior relationship

3 Data Analysis in Dental Sector

Data analysis is a science that uncovers patterns. Machine Learning, or Deep Learning techniques are used to analyze data. The application of Deep Learning techniques requires a large amount of data, and makes it difficult to apply to the current study. The areas to which these techniques can be applied are diverse, ranging from energy optimization and industry 4.0 processes to education and medical care [9], [6], [22], [10]. Data analysis in the dental sector has been a common practice for decades and it has been used to treat different diseases, and dental and mouth-related problems [28], as well as other general problems [19]. Statistical methodologies have made it possible to extract knowledge for better diagnosis and treatment [4], [8].

Initial studies obtained data from questionnaires that were answered by patients [1], [13]. Collecting a large amount of information from patients was difficult and there was a possibility that the patients' answers were imprecise or misleading. Consequently, the results of those studies were not entirely reliable. Fortunately, recent advances in computer technology have simplified the process of collecting data from patients. Moreover, current tools can accurately capture and store information about patients without having to rely on patients' opinions.

Similarly, data processing takes much less time; the advances in computing have increased the processing speed and the ability to apply artificial intelligence algorithms. Therefore, there is a tendency to use this type of methodology in the dental sector.

This is why machine learning methodologies have become more prevalent in the dental sector and have been applied in numerous studies in recent years. For example, machine learning-based methodologies have been employed in the development of a recommendation system for dental care, [15], dental age estimation [26] or for the study of the factors influencing the prognosis of a dental implant [11], [18].

Among the multiple technologies and methodologies for data analysis, a specific subset has finally been chosen and used in this study. This subset combines statistical methodologies with unsupervised and supervised learning methodologies. More specifically, mutual information [5], correlation analysis [12], kernel density estimation [24], association rules [14], and decision trees [20].

4 Analysis

The sample comprised 303 subjects aged between 12 and 15 years, of which 136 were men and 167 were women. The dataset was divided into 3 groups by biotypes or facial patrons: 75 with a brachyfacial pattern, 91 with a mesofacial pattern, and 137 with a dolichofacial pattern.

Both patients and their legal representatives (parents in all cases) signed an agreement by which they provided (disinterestedly) the data for this study, and could not disseminate them for other purposes. For this reason, the dataset cannot be made public.

The demographic characteristics of the dataset are represented in Fig. 4. It is justified below that this factor does not have a significant influence on the class (within the range of values in the dataset in the case of age).

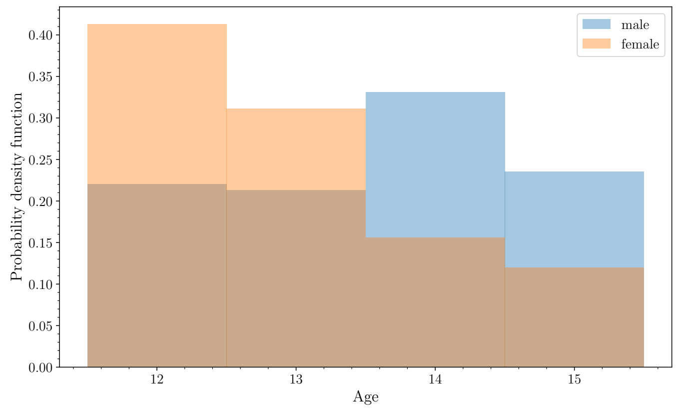


Fig. 4: Gender and age distributions in the dataset.

Fig. 5 shows the values of the mutual information [17, 23] of each of the attributes in the dataset with the target variable, ordered by such values. As can be seen there, age and gender do not have any significant influence on the class, which suggests that the demographic characteristics of the sample (Fig. 4) do not influence the study. On the other hand, it is observed that most of the attributes have a marked relationship with the class, giving rise to the represented ranking, which justifies trying to predict the presence or absence of treatment using a subset of attributes.

The two most relevant features in the mutual information analysis can provide a partial explanation of the presence or absence of treatment. These characteristics were chosen systematically for simplicity of extraction for this first study. This can be observed in Fig. 6, where the distributions for each of the series practically do not overlap.

Since the purpose of the research is to assist in medical diagnosis, and the physician must weigh the importance of the various errors, it is difficult to reduce the performance of the proposed system to a single metric. Instead, several alternatives are offered, as well as the confusion matrix which provides a complete evaluation of the system. However, it is important to realize that the attributes present in the dataset are not independent of each other and that some of them may provide the same information about the target variable. A simple means of analyzing this relationship is by using Pearson's correlation coefficient between the pairs of attributes, as represented in Fig. 7. It is convenient to analyze the correlation of the variables involved in order to identify possibly redudant

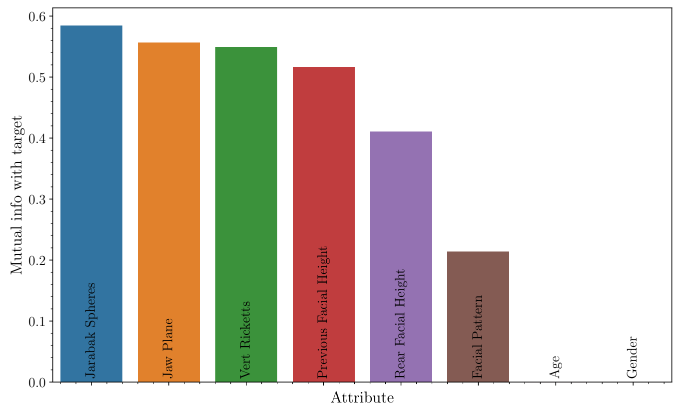


Fig. 5: Mutual information for each of the attributes in the dataset with the presence or absence of treatment.

information, which hinders the performance of the models and makes their interpretation difficult.

It can be seen from the correlation matrix that several strongly related variables are worth studying independently. Fig. 8 illustrates the interrelations between the different variables, as well as the ability to separate the classes they provide in pairs.

Before proceeding to supervised modeling, the dataset has been explored for association rules. Association rules are found by analysing data for frequent if/then patterns, thanks to which it is possible to assess their reliability. The metrics used in the search for association rules include: Support, which measures the frequency with which the consequent of the rule appears when the consequent appears. Confidence, which indicates the number of times both the antecedent and the consequent appear in the dataset. Lift, measures the extent to which rule frequency is greater than it would be in a statistically independent condition; values greater than one indicate an increase from that hypothesis. In this study, the widely used Apriori algorithm, proposed by Agrawal and Srikant in 1994 [2], has been employed for association rules extraction.

To obtain a set of useful rules, the search has been restricted to those with confidence greater than 0.5 and with a lift greater than 1.1, using only categorical attributes. The resulting rules are summarized in Tab. 9. Although it is also possible to extract rules by bucketing continuous attributes, the resulting rules would not provide as much knowledge about the dataset.

As it can be seen in Tab. 9, there are strong relationships between the facial pattern and the class, as well as the Dolicho pattern and the presence of treatment. The other facial patterns are associated with the absence of treatment.

So far, the analysis results show that there are multiple relationships between the attributes and the presence or absence of treatment. To determine, numer-

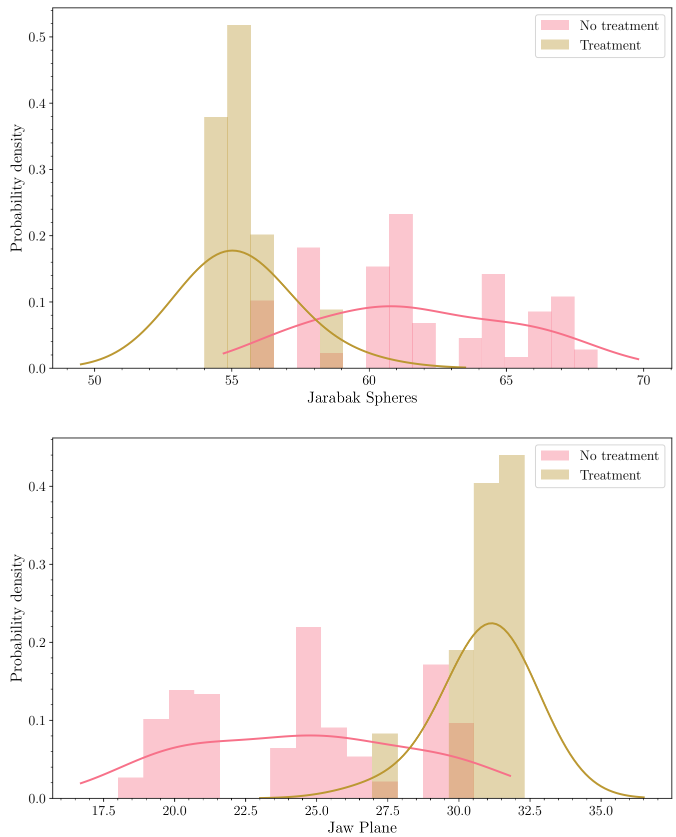


Fig. 6: Probability distributions of the two attributes that are the most relevant to the target variable. The figure shows histograms (shaded regions), as well as the kernel density estimator with manually chosen bandwidths, to illustrate the probability distribution, distinguishing the distributions according to class (legend).

ically, the probability that a patient requires treatment, it is necessary to adequately represent the knowledge. This is difficult because of how the attributes are interrelated. Fortunately, multiple supervised learning algorithms can offer a solution to this problem. Decision trees integrate statistical and mathematical techniques to facilitate the selection of optimal criteria for dataset description. Thanks to its training, the decision trees algorithm can provide human-understandable representations. Fig. 10 shows a decision tree which was fit with the dataset. It can be observed that the limit value of the Vert Rickets variable is -1.25, making it possible to identify the majority of cases in which treatment is necessary. In cases where the value crosses the threshold, the algorithm considers the value of the rear facial height and can identify some of the cases in

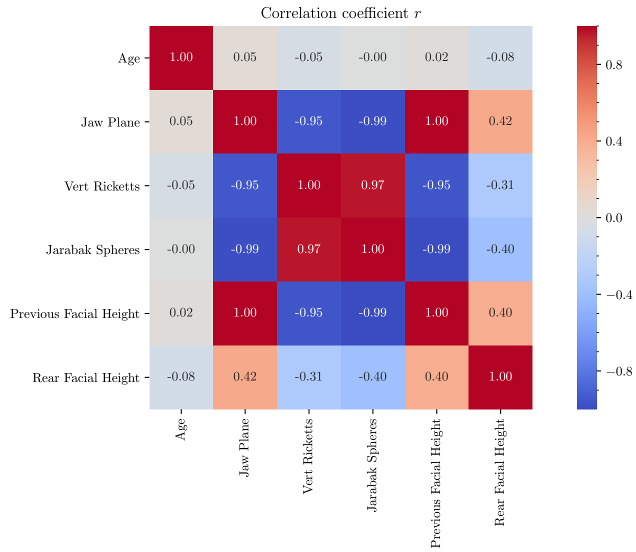


Fig. 7: Pearson correlation coefficients between pairs of variables in the dataset.

which treatment is necessary (less than 70.0, left node). Nevertheless, several limit cases remain in which the algorithm cannot determine whether treatment is necessary or not (more than 70.0, right node).

Despite the limit cases, the tree is a highly optimal model for representing all the data because it only classified 4 patients (1.32 %) erroneously. The actual evaluation of the tree must be made using a procedure which avoids biasing due to evaluating the tree with the training data itself. To prevent this, a 5-fold evaluation has been carried out, the tree has been clinically trained with 4/5 of the dataset and evaluated on the remaining 1/5. The confusion matrix obtained with this method is

$$\begin{pmatrix} 205 & 4 \\ 0 & 94 \end{pmatrix},$$

where rows refer to the actual class (presence and absence of treatment) and columns refer to the prediction of the tree (in the same order).

It can be seen that the evaluation results are consistent with those of the tree shown in Fig. 10. Thus, the algorithm's "failure rate" is estimated to be approximately 1.32 %. Failure occurs when the algorithm misclassifies the patients, suggesting that there is no need for treatment while in reality, an expert would have recommended treatment.

Another possible interpretation of the misclassified instances is that the "impure" leaf node in Fig. 10 (bottom right) actually represents cases where the decision is not so clear and that both results may be acceptable in the expert's opinion.

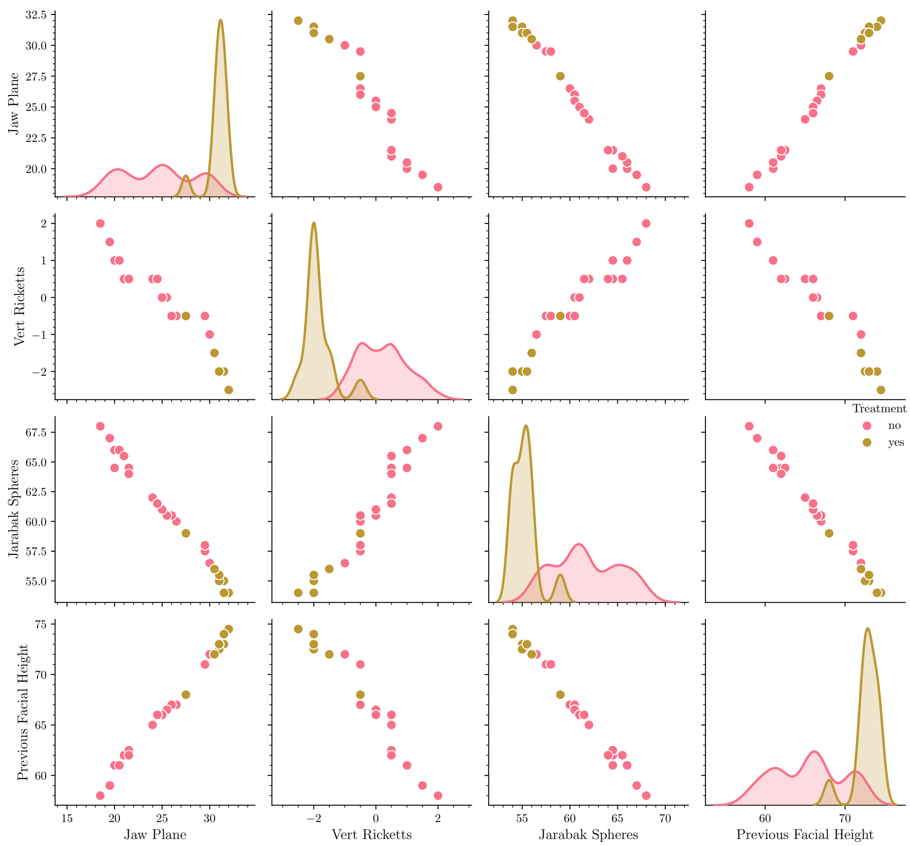


Fig. 8: Representation of variables in pairs in relation to class (color). The plots in the diagonal of the grid represent the distributions using a kernel density estimator.

	antecedent	consequent	support	confidence	lift
2	Facial Pattern = dolicho	Treatment = yes	0.287129	0.635036	2.046979
3	Treatment = yes	Facial Pattern = dolicho	0.287129	0.925532	2.046979
1	Facial Pattern = brachy	Treatment = no	0.247525	1.000000	1.449761
4	Facial Pattern = meso	Treatment = no	0.277228	0.923077	1.338241
0	Facial Pattern = brachy	Gender = male	0.125413	0.506667	1.128824

Fig. 9: Association rules extracted by the Apriori algorithm [2] using the criteria detailed in the main text (lift > 1.1, confidence > 0.5).

5 Conclusions

The purpose of this study was to investigate whether there is a correlation between the craniofacial morphology and the need for tooth extraction as part

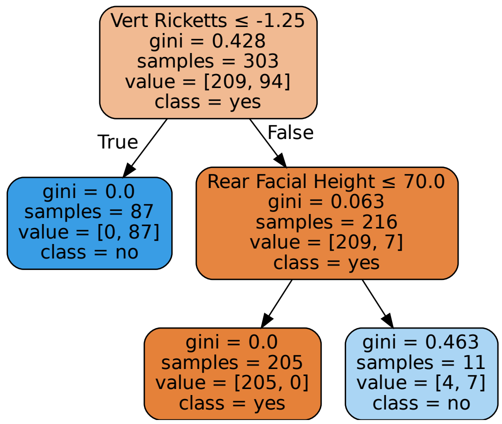


Fig. 10: Decision tree trained with the whole dataset. The first line of a non-leaf node indicates the decision criterion. The gini value describes the tree impurity (Gini impurity). All other information in the nodes can be interpreted directly.

of orthodontic treatment. According to our findings, non-extractive therapy is preferable. This is because the extraction of premolars tends to rotate the jaw counter-clockwise due to the closure of post-wedge mesializing molars. Counter-clockwise rotation makes the chin project forward which worsens the appearance of the profile; it also tends to close the bite, to reduce the lower face height, and to retract the lip protrusion.

Our analysis has shown that there are multiple relationships between the attributes and the presence or absence of treatment. To determine, numerically, the probability that a patient requires treatment, it is necessary to adequately represent the knowledge. This is difficult because of how the attributes are interrelated. However, the analysis has shown that it is easy to control the depth of the bite and the incisive and lip protrusion in patients with a mesofacial pattern. Thus, it is possible to have good treatment control with or without dental extractions because in these patients. Machine Learning techniques have been applied to identify the relationship between craniofacial parameters and the need for tooth extraction. By applying simple rules, it has been possible to identify the need of treatment in 98.7 % of the cases, while the remaining can be regarded as “limit cases”, which must be evaluated by a human expert.

Acknowledgement

The work of José Machado and Paulo Novais has been supported by FCT – Fundação para a Ciência e Tecnologia within the RD Units Project Scope: UIDB/00319/2020

References

1. Agerberg, G., Carlsson, G.E.: Chewing ability in relation to dental and general health: analyses of data obtained from a questionnaire. Acta Odontologica Scandinavica **39**(3), 147–153 (1981)
2. Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB. vol. 1215, pp. 487–499 (1994)
3. Andrews, L.: The six elements of orofacial harmony. Andrews J **1**, 13–22 (2000)
4. Bearden, E., Robinson, K., Deis, M.: A statistical analysis of dental hygiene students' grades in online and on-campus courses and performance on the national board dental hygiene exams. Journal of dental hygiene: JDH **76**(3), 213–217 (2002)
5. Bennasar, M., Hicks, Y., Setchi, R.: Feature selection using joint mutual information maximisation. Expert Systems with Applications **42**(22), 8520–8532 (2015)
6. Candanedo, I.S., Nieves, E.H., González, S.R., Martín, M.T.S., Briones, A.G.: Machine learning predictive model for industry 4.0. In: International Conference on Knowledge Management in Organizations. pp. 501–510. Springer (2018)
7. Dimberg, L., Arnrup, K., Bondemark, L.: The impact of malocclusion on the quality of life among children and adolescents: a systematic review of quantitative studies. European journal of orthodontics **37**(3), 238–247 (2014)
8. Edgar, H.J.H.: Dentitions, distance, and difficulty: a comparison of two statistical techniques for dental morphological data. Dental Anthropology **17**(2), 55–62 (2004)
9. González-Briones, A., Hernández, G., Corchado, J.M., Omatu, S., Mohamad, M.S.: Machine learning models for electricity consumption forecasting: A review. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). pp. 1–6. IEEE (2019)
10. González-Briones, A., Hernández, G., Pinto, T., Vale, Z., Corchado, J.M.: A review of the main machine learning methods for predicting residential energy consumption. In: 2019 16th International Conference on the European Energy Market (EEM). pp. 1–6. IEEE (2019)
11. Ha, S.R., Park, H.S., Kim, E.H., Kim, H.K., Yang, J.Y., Heo, J., Yeo, I.S.L.: A pilot study using machine learning methods about factors influencing prognosis of dental implants. The journal of advanced prosthodontics **10**(6), 395–400 (2018)
12. Härdle, W.K., Simar, L.: Canonical correlation analysis. In: Applied multivariate statistical analysis, pp. 443–454. Springer (2015)
13. HELÖE, L.A.: Comparison of dental health data obtained from questionnaires, interviews and clinical examination. European Journal of Oral Sciences **80**(6), 495–499 (1972)
14. Heritage, J., McDonald, S., McGarry, K.: Integrating association rules mined from health-care data with ontological information for automated knowledge generation. In: UK Workshop on Computational Intelligence. pp. 3–16. Springer (2017)
15. Hung, M., Xu, J., Lauren, E., Voss, M.W., Rosales, M.N., Su, W., Ruiz-Negrón, B., He, Y., Li, W., Licari, F.W.: Development of a recommender system for dental care using machine learning. SN Applied Sciences **1**(7), 785 (2019)

16. Jarabak, J.R., Fizzell, J.A.: Technique and Treatment with the Light-wire Appliances: Light Differential Forces in Clinical Orthodontics. Mosby (1963)
17. Kraskov, A., Stögbauer, H., Grassberger, P.: Erratum: Estimating mutual information [phys. rev. e 69, 066138 (2004)]. Physical Review E **83**(1), 019903 (2011)
18. Lee, J.H., Kim, D.H., Jeong, S.N., Choi, S.H.: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of dentistry **77**, 106–111 (2018)
19. Locker, D., Shapiro, D., Liddell, A.: Negative dental experiences and their relationship to dental anxiety. Community dental health **13**(2), 86–92 (1996)
20. Ma, L., Destercke, S., Wang, Y.: Online active learning of decision trees with evidential data. Pattern Recognition **52**, 33–45 (2016)
21. Ricketts, R.M.: Planning treatment on the basis of the facial pattern and an estimate of its growth. The Angle Orthodontist **27**(1), 14–37 (1957)
22. Rivas, A., Fraile, J.M., Chamoso, P., González-Briones, A., Rodríguez, S., Corchado, J.M.: Students performance analysis based on machine learning techniques. In: International Workshop on Learning Technology for Education in Cloud. pp. 428–438. Springer (2019)
23. Ross, B.C.: Mutual information between discrete and continuous data sets. PloS one **9**(2), e87357 (2014)
24. Silverman, B.W.: Density estimation for statistics and data analysis. Routledge (2018)
25. Svedström-Oristo, A.L., Pietilä, T., Pietilä, I., Vahlberg, T., Alanen, P., Varrela, J.: Acceptability of dental appearance in a group of finnish 16-to 25-year-olds. The Angle Orthodontist **79**(3), 479–483 (2009)
26. Tao, J., Wang, J., Wang, A., Xie, Z., Wang, Z., Wu, S., Hassanien, A.E., Xiao, K.: Dental age estimation: A machine learning perspective. In: International Conference on Advanced Machine Learning Technologies and Applications. pp. 722–733. Springer (2019)
27. Tweed, C.H.: The frankfort-mandibular plane angle in orthodontic diagnosis, classification, treatment planning, and prognosis. American journal of orthodontics and oral surgery **32**(4), 175–230 (1946)
28. Williams, A., Bower, E., Newton, J.: Research in primary dental care part 6: Data analysis. British dental journal **197**(2), 67 (2004)