Journal of Cleaner Production 176 (2018) 813–826

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A collaborative web service exploiting collective rules and evidence integration to support sustainable orthodontic decisions

Ernesto D'Avanzo [a, *], Vincenzo D'Antò [b], Ambrosina Michelotti [b], Roberto Martina [b], Paola Adinolfi [c], Ada C. Pango Madariaga [b], Roberto Zanoli [d]

[a] Department of Political, Social and Communication Sciences, University of Salerno, Salerno, Italy
[b] Department of Neuroscience, Reproductive Sciences and Oral Sciences, University of Napoli Federico II, Napoli, Italy
[c] Department of Business Sciences & Management of Information Systems, University of Salerno, Salerno, Italy
[d] Fondazione Bruno Kessler, Trento, Italy

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form
8 November 2017
Accepted 13 November 2017
Available online 19 December 2017

Keywords:
Sustainable decision-making
Class I malocclusion
Decision trees
Education for sustainable development
Policy for sustainability
Evidence-based decision-making (EBDM)

ABSTRACT

Despite the growing demand for orthodontic care, a framework to support sustainable orthodontic decision-making is lacking, even if scientific literature offers several attempts to deal with this issue. As well known, dentistry generates solid health residues that include heavy metals and biomedical waste, that asks for a professional duty and a social responsibility of the orthodontist that should transform, more and more, his daily practice to a sustainable one, by adopting environmental oriented measures and, at the same time, cutting the overall costs of his professional performance while keeping the performance standards high. This work aims at filling such a gap in knowledge by proposing a *decision tree* algorithm that, besides increasing the level of agreement *within* and *between* orthodontists, allows for the adoption of a framework of sustainable orthodontic *best practices*, using a *dataset* of 290 randomly selected patients generated from 2011 medical records of patients of the orthodontic School at the University of Napoli "Federico II".

The *best practices* framework, provided as *if-then rules* which can be easily inspected by orthodontists, represents a sustainable model in that it minimizes the time and resources employed for dentistry decision-making, dramatically reduce the environmental impact in terms of waste and use of electric equipment and tools, and increases patient satisfaction by delivering quick and appropriate treatment, thus meeting the economic, environmental and social pillars of *sustainability* in health care.

1. Introduction

Dental malocclusions are highly prevalent pathologies in the population and the increasingly close attention to aesthetic and functional problems has led to a larger demand of orthodontic treatments in recent years (Lin et al., 2016). As shown by a survey of the American Association of Orthodontists (AAO),[1] in 2012 AAO members treated a total of 5,876,000 patients, with a 20% increase compared to 2010. Another survey shows how 75% of adult's subjects surveyed reported an increased sense of self-confidence, while 92% of the whole sample of respondents said they would definitely recommend orthodontic treatment to other adults.

However, despite the growing demand for orthodontic care, a framework to support sustainable orthodontic decision-making is lacking. As known, orthodontic diagnosis is highly energy and resource demanding, with important environmental impact. In fact, it asks for huge electricity demands of electronic dental equipment and copious water requirements; there are environmental effects of biomaterials before, during and after clinical use, the employment of radiation and, last but not least, orthodontic diagnosis and treatment cause the production of unsafe waste such as mercury and other waste material. The column "How it is done" of Table A1, in Appendix A, reports all tools (e.g., Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector, and so forth) and waste material (battery, light, etc.) employed to perform an orthodontic diagnosis that exploits the *skeletal*, *clinical*, *radiographic*, and *personal data* features. In order to reduce the effects of environmental deterioration, many forces have been involved worldwide by

* Corresponding author.
E-mail address: edavanzo@unisa.it (E. D'Avanzo).

[1] The survey, titled "The Economics of Orthodontics," asked members of the AAO in the United States and Canada information about patients they were treating in 2012.

https://doi.org/10.1016/j.jclepro.2017.11.093
0959-6526/

employing sustainability concepts and green solutions in several ways, with a real "call to arms" in order to convert orthodontics from an unsafe to a sustainable practice, by adopting a "green dentistry" (Mulimani, 2017). For instance, one attempts to implement sustainability in healthcare has been done by the United Kingdom NHS that promoted advising papers, set up groups to establish measures and carrying out practices through the Sustainable Development Unit. On 2014 the Sustainable Healthcare Strategy was established, a pan-European initiative aiming at supplying solutions to the sustainability of healthcare: European healthcare systems are more and more required to set down better care with reduced resources. Monash Health, a health service based in Melbourne, Australia, sought to establish a program of disinvestment to improve patient outcomes by removing, reducing or restricting health technologies and clinical practices that were unsafe, ineffective or inefficient.

However, despite the institutional involvement cited above, such sustainable frameworks are not yet applicable, particularly in the orthodontics area. Moreover, scientific literature is still free from studies in this field, except for the abundance of works that are predominantly narrative (Pithon et al., 2017).

It is a shared view that, in order to address the problem of health care efficiently and sustainably, it is necessary to study in detail the processes concerning the treatment of patients in different medical conditions, trying to identify the most satisficing possible organization, in terms of resources combination, for each diagnostic-therapeutic pathway.

For instance, a number of authors claim that physicians should exhibit a *sustainable decision-making* because of the scarcity of resources. In this sense, Bodemer et al. (2015) suggest that a *sustainable decision* mechanism should exhibit both high *sensitivity* (i.e., correctly allocating patients requesting specialized care) and low *false positive rate* (i.e., avoiding unnecessary allocation of patients in specialized department if specialized medical treatments are not required).

Scientific literature concerning dental research and practice is rich of studies that pursue the goal of identifying the clinical reasoning of the specialist physician, which translates clinical records into coded choices, and shared *actions/policies* (Musen et al., 2014).

The spread of ineffective and inappropriate treatments has given rise to the development and dissemination of *evidence-based medicine*. Straus and Sackett (1998) proposed a conceptualization of Evidence-based-medicine according to decisions are the result of the integration between the doctor's experience and the conscientious, explicit, and judicious use of the best available scientific evidence, such as diagnostic tests, prognostic factors, effectiveness and safety of preventative treatments, and so on, that, as a whole, are mediated by the patient's preferences. Patient mediation and participation in the decision helped to name this approach *shared decision-making* to indicate that physicians and patients decide on the basis of the *best available evidence* in a sustainable manner (Stiggelbout et al., 2012), for instance they introduce sustainability into a health system by bringing clinical, financial and operational data together to analyze resource utilization and productivity.

Another relevant issue of orthodontic care is represented by the difficulty to make orthodontic diagnosis, due to the *subjective* interpretation of diagnostic records: Kravitz and Bowman (2016), demonstrated that a minimal configuration of a record set for orthodontic diagnosis and treatment planning could not be defined (, 2016). Ribarevski et al. (1996), in their investigation, demonstrated that the level of agreement for the *extraction/not-extraction* decision *within*[2] orthodontists is moderate, and a poor agreement *between* the orthodontists does exist. More recent investigations show that this trend concerning poor-moderate agreement *within* and *between* orthodontists, still holds (Hu et al., 2015). These findings show the *subjective* aspects of orthodontic diagnoses, the lack of *universality* and *unanimity* in the interpretation of orthodontic data and, consequently, in the choice of treatment as claimed by Nouri et al. (2016), suggesting that treatment planning is derived from weak levels of scientific evidence (Turpin and Huang, 2016).

On the whole, the above evidence shows that a referencing framework for a sustainable orthodontic decision-making would be desirable and beneficial for a diagnostics treatment selection, particularly as regards controversial cases, where *subjective* data interpretation could generate unappropriate decisions (Nguyen and Proffit, 2016). Such a framework would be particularly useful to improve the sustainability of the care provided.

In this sense, innovation plays a chief role in enhancing sustainability and represents a key area confronted by the sustainable development discourse (Matos and Silvestre, 2013), through which public and private organizations can accomplish change and, at the same time, turn more sustainable (Silvestre, 2015). Patients, however, can benefit from innovation only if it is affordable now and sustainable in the future.

This paper introduces a framework to identify *best practice* in the form of *rules*, automatically generated by a *decision tree* algorithm, that, besides increasing the level of agreement *within* and *between* orthodontists, allows for the adoption of a sustainable orthodontic practice. It integrates the three main pillars of sustainability (economic, environmental and social), increasing efficiency, minimizing pollution and improving quality and patient satisfaction in the day-to-day practice.

2. Literature review

Decision tree is a classification scheme that generates a *tree* and a set of *rules* from a given dataset (Witten and Frank, 2011). It has been widely employed both to represent and run decision processes (Anderson et al., 2015). Considering that medical decisions are made for various purposes including screening, diagnosing, and treatment prescription, the decision problem becomes difficult to visualize and implement (Croskerry, 2015). A *decision tree* represents a useful graphical tool in such settings, as it allows for intuitive understanding about the problem and can aid decision-making since it is interpretable through *if-then rules* by any orthodontist, even if the physician is not trained in computer applications. For instance, Table 8 shows just a set of these kind of *rules* generated by the *decision tree* in Fig. 4. Any orthodontist, even trainee, could refer to such a kind of rule in order to take a treatment decision on the basis of a very short ordered list of *features* (i.e., *attributes*[3]).

The approach introduced in the following pages represents a

[2] "Within agreement" is a jargon expression that indicates the level of agreement that orthodontist *O* has with his treatment decisions over time, compared to the same patient P. For example, in time t_1, physician *O* might have decided on extraction treatment (or non-extractive) relative to the tooth x of patient P, while in t_2 time it could opt for non-extractive (or extractive) treatment for the same tooth x of patient P. In this case, the physician has a "within agreement" 0 for patient *P*. If this fluctuation in the decisions of the same doctor occurs for several patients $P_1 \ldots P_n$, it is said that the rate of "within agreement" is low for doctors *O*. A similar argument applies to "between agreement". Only for the latter, the *x* rated agreement regarding patient P is no longer the same doctor over time but a team of doctors, one with respect to the other at the same time t.

[3] In the continuation you use indifferently *attributes* and *features*.

further attempt, along with the others, towards the foundation of a common framework aimed at reducing, as much as possible, *subjectivity* in the interpretation of orthodontic data (Masías et al., 2015). For instance, Sammut and Webb (2011) employ *neural networks* for teeth extractions decision-making. To this end, the authors employed a dataset of 156 patients made of 12 cephalometric *features* and 6 additional *features*. Extraction patterns, exploitable for orthodontic treatment decisions, were obtained applying four *neural networks* that make use of a back-propagation algorithm. Experimental results show that success rates of the models generated was 84% for the decision of *extraction* vs. *not-extraction*. Xie et al. (2010) applied a *neural networks* for the orthodontic treatment of patients aged 11–15 years old, in order to determine extraction treatment. Experimental settings employed a dataset made of 200 subjects, using 23 features. The experiments allowed for estimating the contributions of the 23 input *features* to the final output (i.e., *extraction* vs. *not-extraction*). For instance, "Anterior teeth uncovered by incompetent lips" and "IMPA (L1-MP)" resulted to be the two *features* that give the biggest contributions sequentially. According to the authors, when clinicians are predicting whether an orthodontic treatment requires extraction, the features "anterior teeth uncovered by incompetent lips" and "IMPA (L1-MP)" should be taken into consideration first. Martina et al. (2004) developed a decision support system based on *neural networks* in order to aid clinical decision-making for orthodontic extractions. The employed *neural network* makes use of a feed-forward back-propagation paradigm trained on a dataset made of 48 cases, exploiting, overall, 32 *cephalometric* and orthodontic *cast* measurements as *features*. As for the evaluation, the system output was considered correct if its decision (i.e. *extraction* or *not-extraction*) coincided with the decision for the patient at the moment of the orthodontic treatment. In both cases, the performance of the system achieved an accuracy level of 75%.

Although *neural networks* outperform *decision trees* for some tasks, however, they need large amount of annotated data that are not always available, just like in this work. Then, *neural networks* result slower, both for *training* and *classification*, and are not suitable for real-time web services such as the one proposed in this experimental framework. In addition, *neural networks* generate unintelligible models that are not designed to be interpreted by humans who can only see the output of the process without the possibility of exploring it. Instead, orthodontists that employ *decision trees* have at their disposal models that can be inspected as *paths* on *trees* as well as in the form *if-then rules* such as those reported in Table 8; *rules*, or the corresponding *paths* on the *tree,* are comparable with *best practices* that physicians adopt in their professional activity. Moreover, *rules* generated by the *trees* can be easily employed by *trainees* at the Orthodontic School that can benefit by learning and employing *objective* sustainable *best practices* that have been preventively shared and validated by senior physicians. Then, once the model has been validated, classification performance can always be improved, for example by learning multiple *trees* of different subsets of the training, by employing *AdaBoost* algorithm, that can be applied concurrently with several *decision tree* learning algorithms, with the former aiming at improving individual performance of the latter. Another major problem emerging from the scientific literature introduced above, concerns evaluations proposed for the algorithm employed. In fact, all the experiments performed have always employed evaluation *metrics* that return an *overall accuracy*, without considering specific performance (i.e., *detailed accuracy*) for individual classes, in which case the performance of the algorithm would be expected to fall. Moreover, the number of training examples used to train *neural networks* introduced above is too small, since the practice suggests having about 10 examples for each *feature* (Figueroa et al., 2012). Otherwise, the learning algorithm, whatever it is (e.g., *neural network* or *decision tree*), would suffer from *overfitting*; in other words it turns to poor predictive performance because it reacts to all, even minor, fluctuations in the training data and, since the learning model depends on it, it is likely to have a higher error rate for new unseen data. But even more important is the question of a missing clinical evaluation for all works reported above, since all the proposed approaches aim to support the physician in his clinical decisions, on a daily basis.

Thanks to the methodology introduced in Section 3, physicians, who first exhibited low *within* and *between agreement*, showed an increase in the agreement levels when a set of candidate *best practices* was given to them, as *if-then rules* automatically generated by the *decision tree*, such as those in Table 8. This demonstrates the benefit of employing the approach proposed in terms of quality of clinical decisions, which impacts on the level of patient satisfaction, thus corroborating the social pillar of sustainability.

3. Material and methods

From a database of 2011 medical records of patients, collected at the University of Napoli "Federico II" in the last two years, it has been generated a *dataset* of 290 randomly selected patients' data that are representative of the underlying population of interest. Extracted patients aged 8 to 53 (with a *mean* of 15.59 years of age, and a *standard deviation* of 5.99), in the permanent dentition, without previous orthodontic intervention. The average age is that of a young population, since in most cases a treatment is given to 12 years of age. The adult population, albeit increasing, is always a minority, as evidenced by scientific literature reported in Section 1. Fig. A2, in Appendix A, shows the frequency distribution of patients by age.

Each scholar/practitioner at the Orthodontic School contributed to builds his own *subset* of the whole *dataset* by detecting, for each patient, 39 common *features*, that are adopted by all practitioners in order to describe the case, including the *class label*, (teeth *extraction* or *not-extraction).* Subject's *features* were divided according to *skeletal class*, *clinical*, *radiographic*, and *functional* features. Table A1 in Appendix A reports a complete classification of the *features* employed. The *dataset* counts 232 *negative* examples, that is medical records of patients classified as *not-extraction* cases, and 58 *positive* examples, that is medical records of patients classified as *extraction* cases. The *dataset* shows a situation of unbalanced data distribution (i.e., *skewed*, in statistical terms) with respect to the *class label/target value* (i.e. teeth *extraction* or *not-extraction*) that needs to be modeled. This situation will require certain steps during the training of the algorithm, in order to take account of the lower weight played by the *positive* examples, within the entire economy of the *dataset*. The experimentation, and the testing, delivered through the web service, as already said, is designed to support treatment options for *Class I malocclusion*.

Experiments introduced below, and the corresponding web service prototype based on them (available at www.coltho.org), employ J48 *decision tree*, a WEKA implementation of C4.5 *decision tree* algorithm, developed by Quinlan (1993). C4.5 *decision tree* classifies instances, i.e., orthodontic medical records, by sorting them down from the *root* to some *leaf* nodes, providing the classification of the instances (i.e., *extraction* $= 1$ vs. *not-extraction* $= 0$). Nodes of the *decision trees* specify tests of some *features* describing the instances, such as *goGnLi* at the root node of the *decision tree* in Fig. 4. *Branches* descending from nodes correspond to one of the possible values the attribute may assume; for instance, *goGnLi* in Fig. 4 may assume two sets of possible values, those $<= -104.2$ and those > -104.2. The same process is repeated for the *sub-tree* rooted at the new node. Looking at Fig. 4, after testing *goGnLi* at the root node, J48 jumps on the right and left branches, based on the two sets of value the *root feature* may assume, and tests, respectively, *apgB1* and *snpg*. The process is repeated until a leaf node is reached, where the class label is present (*1*or *0*).

The *feature selection*, i.e., which feature is to be tested at each node of the tree, plays a chief role for *decision tree* construction. In the experiments introduced below, two *feature selection* methods have been employed, *Information Gain* and *GainRatio,* as masterfully reported in Mitchell (1997). *InfoGain* is strictly related to *Entropy* (Mitchell, 1997), an index of the purity of a *dataset*, since it just represents the expected reduction in *entropy* that results from the partition of the examples according to this *attribute*. For instance, for the orthodontic dataset, *Entropy* is about *0.22*, a value that indicates an unbalanced distribution towards the *not-extraction target*/*class label* (i.e., 0). A drawback of *InfoGain*, however, is that it tends to prefer attributes with many values. *GainRatio* is a possible remedy to this issue, since it levels the playing field by penalizing the multiple-valued attributes (Mitchell, 1997). Fig. B2 in Appendix B reports the *ranked list* of *features*, obtained from the orthodontic dataset, after employing the *InfoGain* and *GainRatio* mechanisms for experiment one described in Section 3. Fig C1 and D2, respectively in Appendix C and D, show a comparison between the two feature selection strategies for experiments two and three, also detailed in Section 3.

Experiments performed have been tested using different evaluation metrics (Fawcett, 2006). As first evaluation metric, *accuracy* has been employed. It measures how often *decision tree* makes the correct prediction, calculating the *ratio* between the number of *correct predictions* and the *total number of predictions*. However, this metrics presents an important drawback since it does not make distinction between classes; correct answers for each class are treated equally. As mentioned in Section 2, this drawback is present in many scientific works, such as those outlined above, which, though apparently having higher *accuracy* values, actually treat classes equally, even when they are not equivalent. For instance, how many examples failed for each class? This is the case for the skewed distribution for the orthodontic *dataset*. Furthermore, it does not distinguish those cases in which the patient underwent an *extraction* when, instead, should not have (i.e., *false positive*), or other cases where a necessary extraction was not performed (i.e., a *false negative*). For such a kind of evaluation the *confusion matrix* was employed, showing a detailed breakdown of *correct* and *incorrect* classifications for each class; such type of information would otherwise be lost just looking at the overall *accuracy*.

Precision score estimates how many cases are actually needed so that the *decision tree* assigns an extraction target, while *recall* allows for determining how many cases are found to be true by the *decision tree*, out of all the cases that are true. *Precision* and *recall* can be read in the light of their *harmonic mean*, F_1, which, unlike the arithmetic mean, tends toward the smaller of the two elements, resulting small if either *precision* or *recall* is small.

In machine learning terms a *learning curve* represents the generalization performance of the model as a function of the size of the training set (Sammut and Webb, 2011); in other words, it depicts improvement in performance on the vertical axis when there are changes in another parameter (on the horizontal axis), i.e. the training set size.

Clinical validation plays a chief role in this experimentation and has a twofold objective. First, it aims to measure decisions *within* and *between* orthodontists where, as demonstrated by the previous work reported in Section 1, the former seems to be poor/moderate, whereas the latter resulted to be poor. Second, the clinical control aims at validating the automatically generated models. Since orthodontists' decisions, as mentioned before, suffer from lack of objectivity, *decision trees* aim to support them with *trees*/*rules* that must be shared as much as possible among physicians who have a low rate of agreement, with themselves (i.e., *within*) and between them (i.e., *between*).

This motivates more strongly our choice to adopt *decision trees* that, as said before, in Section 1, are able to produce inspectable models, unlike other classifiers such as support vector machines or neural networks. Last but not least, rules allow for a reduction in the cost of the treatment and in the waste material.

4. Experiments, results and discussion

Three runs of experiments were performed, in order to generate different *decision trees* to be submitted to medical validation and, therefore, to test their efficiency from a clinical point of view. All details about the results of the experiments are reported, respectively, in Appendix B, Appendix C and Appendix D. Each Appendix, apart from the first one, contains the *decision tree* generated for the experiment and a table reporting the *attribute ranked list* that compares attributes selected through *InfoGain* and *GainRatio*. The *attribute ranked list*, by itself, already allows clinicians to have all the features available, as proposed by the work of Xie et al. (2010) reported in Section 1 that, however, has only two features and no order of importance among them, as the *neural network* return an output without ranking. The approach proposed in our paper, instead, in addition to returning a *ranking* of all *features*, also offers a way to read them, either through a *path* on the *tree* or the corresponding *if-then rule*, representing, as a whole, an added value compared to what has been proposed in the literature. Furthermore with respect to all approaches reviewed in Section 1, the *decision tree* outperforms them also in terms of classical metrics, as shown in the following pages.

To build the *decision tree* model out of the orthodontic dataset, for all the experiments performed, J48 has been run employing *Leave-One-Out Cross-Validation* (LOOCV), in order to estimate the generalization capability of models created by the *tree*, rather than of the model itself (Mitchell, 1997). LOOCV estimates the generalization performance of a model trained on n-1 samples of data, which is a pessimistic estimate of the performance of a model trained on n samples. In other words, to train the *tree*, and avoid the over-fitting problems mentioned above, the most unfavorable condition was preferred: it is less dependent on the fluctuations of training data, and this represents another advantage of our method.

Using this setting for the first run, J48 generated the *decision tree* shown on Fig. B1 (Appendix B), performing with an *overall accuracy* of 71.37%, as reported in Table 1. To get a better overview for both *extraction* (i.e., 1) and *not-extraction* (i.e., 0) class labels, Table 2 shows a detailed *accuracy*, for each class, in terms of *precision*, *recall* and F_1. Even if these results could appear satisficing with respect to the basic setting, since they outperform the performance of the *baseline* (i.e., a *trivial acceptor* whose *precision*, *recall* and F_1 were, respectively, 20%, 100% and 33%), two “singularities” were noted on the learning curve, as can be seen in Fig. 1. In fact, although it showed an increasing trend, two local falls were visible, in terms of performance, towards 50% and 75% of the model training. This behavior prompted us to employ a new experimental setting. However, before doing this, J48 was tested for its capability to address the *skewed distribution* of the orthodontic dataset towards the *not-extraction* class. This is a non-trivial task from a clinical point of view, for two reasons. First of all because in recent years a common trend has emerged: treating patients with orthodontic appliances, rather than resorting to *extraction*. So, as new medical records are collected to update *decision trees* models, a greater number of *not-extraction* compared to *extraction* cases is expected. Secondly, *extraction* treatment planning is a decisive and critical moment for the clinician, since it is a non-reversible procedure, and

Table 1
Overall accuracy for the first experiment expressed in %.

Classification Summary		
Correctly Classified Instances	207	71.34
Incorrectly Classified Instances	83	28.62
Total Number of Instances	xx	

Table 2
Detailed accuracy for the first experiment expressed in %.

	Precision	Recall	F_1
Weighted Average	70.8	71.4	71.1
Class: 0	81.7	82.8	82.2
Class: 1	27.3	25.9	26.5

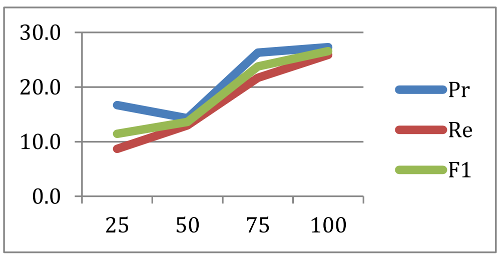

Fig. 1. Graph of the learning curve for the first experiment.

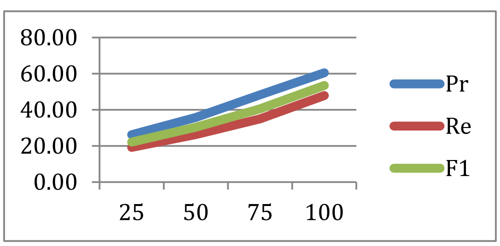

Fig. 2. Graph of the learning curve for the second experiment.

Table 3
Detailed accuracy for the first experiment, but with SMOTE, expressed in %.

	Precision	Recall	F_1
Weighted Average	72.8	73.1	72.9
Class: 0	82.9	83.6	83.3
Class: 1	32.1	31.0	31.6

it is most of the time based on the practitioner's experience. For these reasons, it is critical to provide J48 with a mechanism capable of balancing the distribution towards the *extraction* cases and give them a greater characterization. To this end, we preferred to choose the first configuration of J48, so as to get in the most unfavorable situation. In order to resample the dataset, we employed the SMOTE[4] methodology (Chawla et al., 2002; Weiss, 2004). Results of the experiments are reported in Table 3. As you can see, there is a general and noticeable improvement, both in terms of *weighted average* and for the individual classes, demonstrating the feasibility of the approach also for better settings of J48.

As for the second experiment, we requested J48 to return all “difficult cases”, i.e. those that could be difficult or impossible to classify. J48 returned a list of 15 *medical records*. Among the reported records, 5 cases fell under the *not-extraction* class, while 10 under the *extraction* class. With this new configuration, J48 showed an overall improvement of all its performance. As reported in Table 4, the *overall accuracy* increased from 71.34% of the first setting to 85.45% of the new one. The *detailed accuracy*, per class, also confirms the improvement, as shown in Table 5. This table shows a general increase in terms of the *weighted average* F_1, from 71.1% of the previous configuration to 84.8% of the new one. Looking at the individual classes, J48 exhibits an improvement of the same order of magnitude: F_1 for the *extraction* class (i.e., 1) increases from 26.5% of the previous setting to 53.5% of the new one, whereas for the *not-extraction* (i.e., 0) class F1 increased from 82.2 to 91.4%. As for the learning curves, Fig. 2 clearly shows how the singularities emerged in the previous configuration, have disappeared, showing an increasing linearity of the learning curve that suggests a certain reliability in improving the model, and in generalization, when new medical records appear. All metrics reported above show how *decision tree* outperform models alternative methods.

This setting represents the first intersection point between the automatic evaluation of the generated *decision trees* and the clinical evaluation introduced shortly after. In fact, the *decision tree* generated at this step, showed in Fig. C1, has been dispensed to the members of the Orthodontics School in order to evaluate its *soundness* and *efficiency*. Another intersection point between this configuration setting and clinical evaluation is provided by the “difficult cases” that were submitted to the physicians in order to test their difficulties to classify them. This second trial allows also for evaluating orthodontist *within* and *between agreements*. However, before introducing the clinical evaluation, it is interesting to watch at the last configuration setting of J48. The third configuration was set exploiting C4.5 inner capability of post-pruning (Mitchell, 1997). The technique allows for reducing the size of the *tree*. As for the orthodontic decision-making problem, it allows to provide clinicians with a different visualization of the *tree*, as well as with little change in the rules generated, and the same performance. The third *tree* generated after the pruning step is showed in Fig. D1. As can be noted, unlike the *tree* generated in the second configuration, this one comes in a “lean” and “elongated” shape. The performance of the new *tree* does not change; at least it does not change in terms of *overall* and *detailed accuracy*, as demonstrated by the results reported in Tables 6 and 7. Only a slight decrease in the learning curve is to recorded, however, it continues to grow with linearity and no drop points, as detailed in Fig. 3.

As for the first experiment of clinical evaluation, in order to obtain the classification of the 15 “difficult cases”, 20 orthodontists were given a blinded *excel* table with all 15 medical records, i.e. where record IDs had been removed as well as the earlier classification (i.e., *extracted* or *not-extracted*). During the administration of the test, physicians were told to classify the records, generated randomly from a *decision tree*, looking only at 38 out of the 39 features that describe the case (*class labels* were removed). In the first round, a group of 10 orthodontists annotated 15 medical records. In the second round, when the annotation was blind, only 4 records registered the same annotation of the first round, spread on 4 different physicians from the first round. On the whole, these results show a *within* moderate/poor *agreement* (i.e., about 26%), in line with the scientific literature mentioned in section 1 (Hu et al., 2015). Later, the agreement *between* annotation/physicians was tested looking at how other physicians shared the decision that their colleague made in the first round. For a patient belonging to *not-extraction* class, the decision was shared by 19 colleagues; for another patient, always belonging to *not-extraction,* the decision was shared by 18 colleagues; for one patient belonging to class 1

[4] SMOTE is the acronym of Synthetic Minority Over-sampling Technique.

Table 4
Overall accuracy of the second experiment expressed in %.

Classification Summary		
Correctly Classified Instances	235	85.45
Incorrectly Classified Instances	40	14.54
Total Number of Instances	275	

Table 5
Detailed accuracy of the second experiment expressed in %.

	Precision	Recall	F_1
Weighted Average	84.4	85.5	84.8
Class: 0	89.5	93.4	91.4
Class: 1	60.5	47.9	53.5

Table 6
Overall accuracy of the third experiment expressed in %.

Correctly Classified Instances	235	85.4545
Incorrectly Classified Instances	40	14.5455
Total Number of Instances	275	

Table 7
Detailed accuracy of the third experiment expressed in %.

	Precision	Recall	F_1
Weighted Average	84.4	85.5	84.8
Class: 0	89.5	93.4	91.4
Class: 1	60.5	47.9	53.5

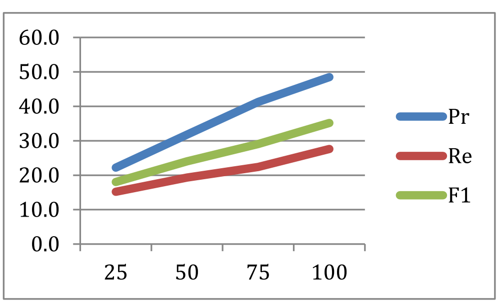

Fig. 3. Graph of the learning curve for the third experiment.

(i.e., *extraction*), the decision was shared by 18 physicians. Seven cases were shared by less than 3 physicians, demonstrating an overall moderate *between agreement*, in line with the previous body of knowledge.

As for the second clinical experiment, *decision trees* generated by J48 in the second and third experimental setting, and showed in Figs. 4 and 5, were submitted to clinicians in order to test their aid in supporting their decisions. During the administration of the test, physicians were required to trace, on the printed *tree*, all possible *paths* (i.e., *best practices*) that were considered useful and plausible to make a positive decision (i.e., *extraction*), even if the *path* ended in a *leaf* node with 0 (i.e., *not-extraction*), so as to avoid any *false negative*.

Results for the *tree* reported in Fig. 4 show that only eight rules have been traced. *Rule 2* is the most frequently suggested by physicians. It gets to the leaf node employing only 4 *features* on the upper part of the tree. Also *rule 1*, shared by 9 physicians, employs only 4 *features* out of 38 to make a decision. Five physicians share *rules 3* and *4*. Whereas the former uses 4 *features*, the latter takes only 2 *features* to make a decision. Only one *path* (i.e., *rule 6*) employed more than 6 features but only one physician suggested it.

Turning to the *decision tree* generated with the third setting, and shown in Fig. 5, three main *rules* were considered plausible by physicians. *Rule 1*, in red, is the most frequent (12 clinicians suggested it) and employs only 3 *features* located in the upper side of the *tree*. *Rule 2* is suggested by 10 physicians and takes a longer *path* (5 *features*). Finally, *rule 3* is suggested by 2 physicians and employs 7 *features*. The good news is that all the *rules* chosen by the clinicians, with the exception of a particular case, ended in the positive *class label* (*extraction*) The two clinical evaluations, that earlier appeared to be unrelated to each other, indeed converge towards a common rationale: physicians, that in the first clinical evaluation exhibited low *within* and *between* agreement, when given a set of candidate "best practices" in the form of *if-then rules* show an increase in the level of agreement, demonstrating the benefit of employing such a model to support a sustainable orthodontic decision-making. Almost all the *rules* chosen reported in Fig. 4, except one, are located in the upper side of the *decision tree* and employ a small number of *features* with respect to the initial list of 38 *features*.

As just mentioned above, orthodontists, by preferring the simplest *hypothesis* that fits the *data*, behave just like any other scientist. This disconfirms what is argued by Turpin and Huang (2016), and cited in Section 1, regarding the lack of scientific evidence for orthodontic job, and, most interestingly, the lack of sustainability. The most important result is that all the rules identified and reported in Table 8 show a significant reduction both in terms of cost of the treatment and in terms of solid waste produced. The format could represent the privileged toolbox of the physician to choose the rules to use in his practice without resorting to any computer support. The orthodontists may observe the sustainability of the best practice/rule chosen both in terms of cost and waste reduction (see Table A1 for a full description of waste material).

5. Conclusion

Further improvements are to be carried out.

A more structured annotation will be requested to physicians, using a Likert scale for each extraction decision. The results of this annotation scheme will be processed by a SEM model (Novo-Corti et al., 2015) that will help to identify the role of specific features with respect to a final model of constructs that is conjectured to exist in the orthodontic practice.

Furthermore, the platform will be enriched in order to include data on costs, measured on the basis of Activity-Based-Costing principles, so as to allow for an economic evaluation of the savings produced by the proposed algorithm in economic and environmental term.

Finally, an instructional framework will be developed, through which the elicited model could be taught to novice orthodontists by means of role playing (i.e. active learning), simulating exercises, gaming, e-learning and chatbox, which will make studying easier and more time-efficient, by instantly delivering feedback back to students in response to work-related questions (Wals, 2014).

In the light of all the peculiarities outlined, the proposed collaborative web service represents a significant contribution to the sustainability of healthcare, defined as a balance of economic concerns the needs of patients and environmental costs (Jameton and McGuire, 2002).

First of all, the proposed methodology allows for defining the minimal configuration of a record set of features for orthodontic diagnosis and treatment planning, thus allowing for a reduction of costs for both diagnostic and clinical trials. By incorporating data on

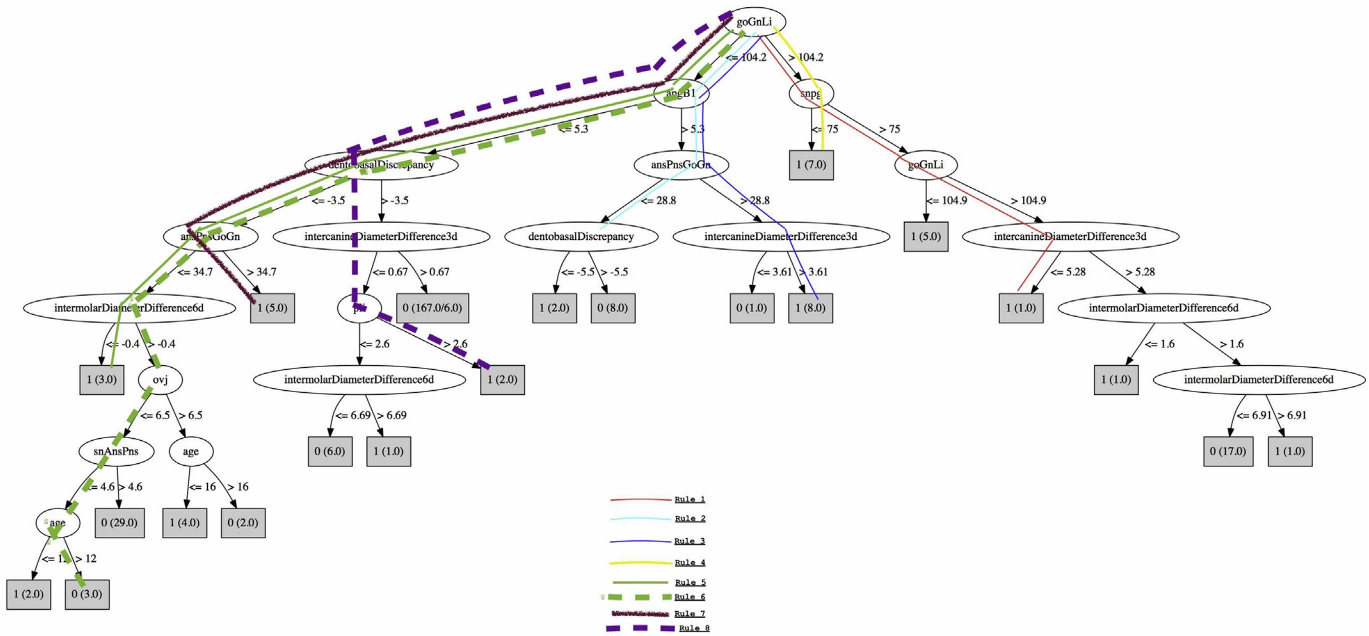

Fig. 4. *Decision tree* generated with the second experimental setting and validated with the rules annotated from physicians.

the costs measured according to activity-based criteria, it may allow to choose the *paths* which are more sustainable, *coeteris paribus*, thus leading to a decrease of the overall cost of clinical treatments and an improvement of organizational efficiency. In different organizational contexts, the same recommended evidence-based *path* may imply different levels of sustainability. So, the proposed algorithm not only reduces the number of *attributes* needed to make a medical decision, but also allows for evaluating their costs, thus contextualizing clinical practice in the specific organizational setting.

Because of its capacity to take into account not only the prevailing scientific references but also the local variables and organizational factors, the proposed methodology can be successfully adopted on large scale by the healthcare systems in a large range of medical procedures. If we adopt a global perspective of sustainable development, it allows transferring knowledge as part of development cooperation with low-income countries.

Furthermore, the reduction of trials allows for containing waste material (water, battery, light, and so on) and the use of equipment and tools (mechanical chair, Nikon 1J5 camera, and so forth), employed to perform an orthodontic diagnosis that exploits the *skeletal*, *clinical*, *radiographic*, and *personal data.* It allows also for a reduction of the risks linked to the use of radiation and to unsafe waste such as mercury. Ultimately, the ecological impact, another pillar of sustainability, is significantly improved.

Indeed, also the social pillar of sustainability is positively affected. The approach goes beyond the simplification of the orthodontic decision-making process, since the models extracted can guarantee greater "objectivity" of the medical decisions in the interpretation of the orthodontic data, and thus allows for the selection of the most appropriate therapy, based on the collective model, previously validated by experts and senior scholars (Hammond et al., 2015). The decision mechanism shows a high level of accuracy, exhibiting both high *sensitivity,* namely low *false negative* (i.e., correctly allocating in specialized departments patients requesting specialized care) and high specificity, namely low *false positive rate* (i.e., reducing unnecessary allocation of patients in specialized departments if specialized medical treatments are not required). As a consequence, patients needs are better served and some aspects which positively impact on patient overall assessment of the performance of a health service and therefore their satisfaction (Faezipour and Ferreira, 2013) are improved, such as treatment length, appropriateness and quality of diagnoses and treatments, perceived competence of physicians, clarity of physicians' communication to patients.

Moreover, the platform is able to deliver its sustainable *rules*, in terms of costs, patient satisfaction and impact on the environment, to scholar and practitioners, providing refresher courses and training for Orthodontic Schools, thus establishing a collaboration among orthodontists and generating learning value (Lozano et al., 2013). This further contributes to the sustainability of healthcare, if we agree with that the sustainable healthcare model should also include the needs of healthcare personnel.

Another peculiar characteristic of the proposed algorithm, which distinguishes it from other methods such as neural networks, is that it allows to trace adherence to the extrapolated model and, by incorporating activity-based costing data, can provide real-time feedback on deviations and possible alternatives: thank to this, a deviation might reveal to be an innovation which improves the sustainability. Besides, unlike neural networks which are slow, unintelligible and require a large number of annotated data (not always available), the proposed procedure can be easily inspected, by means of friendly visualization tools, as *paths* on *trees* or *if-then* rules which are comparable with best practices adopted by physicians in their professional activity. In line with the concept of *shared decision,* the platform is able to benefit from the contribution of its users, who can update the dataset and, as a consequence, refine the model or generate new models. According to sustainable healthcare systems will depend fundamentally on learning mechanisms built into the system to enable continual improvement and adaptation through time. As Heinrichs and Laws (2014) observe "if this goal were to be achieved, a sustainable health care system could be an important step toward a "sustainability state" on a more general level".

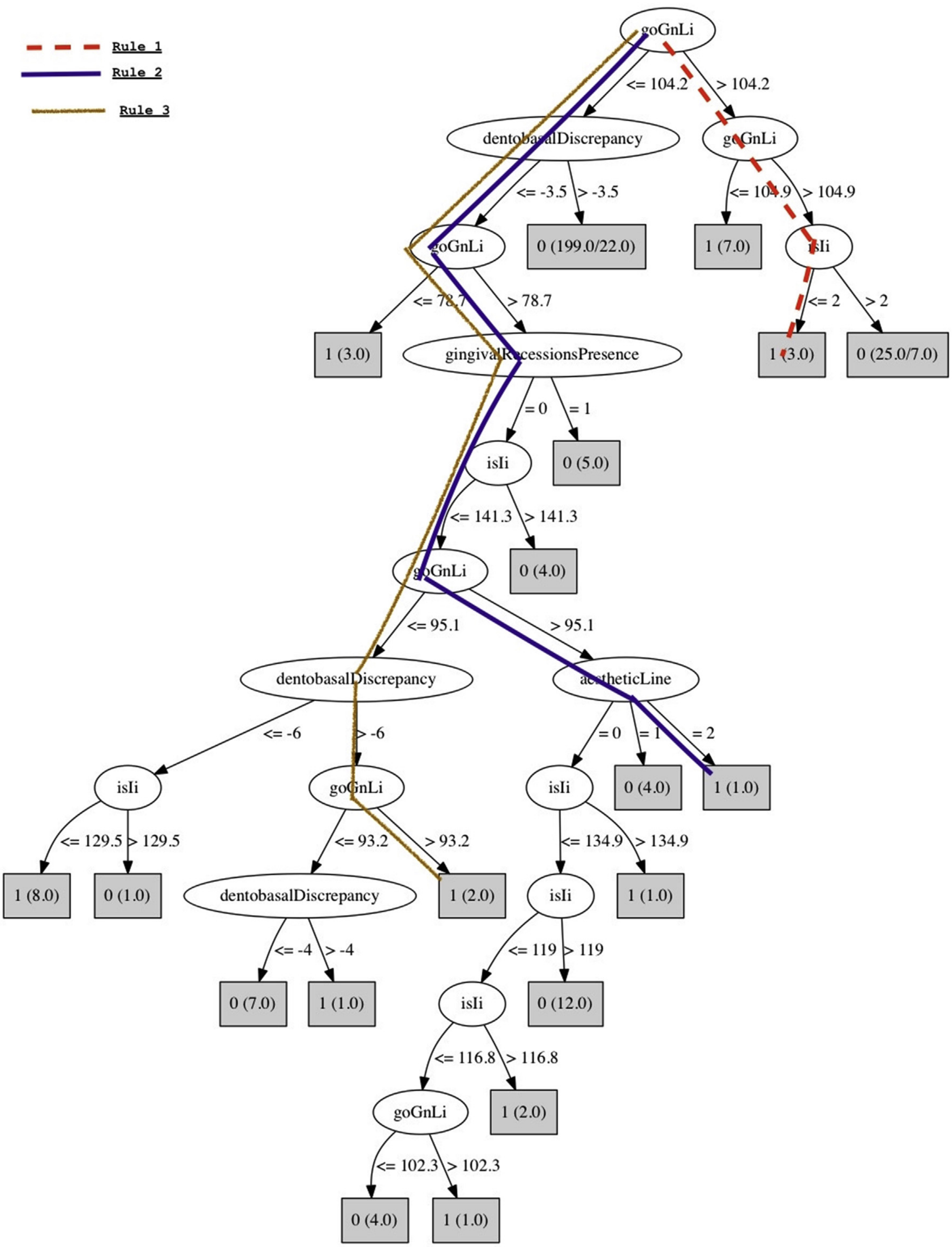

Fig. 5. *Decision tree* generated with the third experimental setting and validated with the rules annotated from physicians.

Table 8
The rules generated by C4.5, readable in the form of if-then rules.

Rule n.	goGnLi	snpg	goGnLi	intercanine diameter difference	apgB1	Ans-Pns^Go-Gn	dentobasal Discrepancy	Class	% of waste reduction (cost/waste)
1	>104,2	>75	>104,9	<=5,28				1	−30/-65
2	<=104,2				>5.3	<=28,8	<=5,5	1	−30/-65
3	<=104,2			>3,61	>5.3	>28,8		1	−30/-65
4	>104,2	<=75						1	−50/75

Appendix A

Table A1
Feature set with names, values and category division of the features (*skeletal*, *clinical*, *radiographic*, and *personal data*) - Nikon J5 camera: Battery / Batteries EN-EL24 Lithium-ion Battery, Battery Life (shots per charge) 250 shots (CIPA), Movies: Approx. 60 min (1080/30p), AC Adapter EH-5b AC Adapter Requires EP-5F Power Supply Connector (available separately).

ID	Feature category	Feature full name	Feature name/ acronym	Feature value	How it is done (tools and material employed)
1	Personal data	Age	Age	Years	Nothing
2	Personal data	Sex	Sex	Binary (Male = 1, Female = 0)	Nothing
3	Clinical	Dentobasal discrepancy	Dentobasal discrepancy (DBD)	Millimeters	Digital Caliper (0–150mm/0–6″ Large LCD Digital Display Vernier Caliper)
4	Clinical (assessment models cass)	Intercanine diameter difference	Intercanine diameter difference (3 diameter)	Milimetres	Digital Caliper (0–150mm/0–6″ Large LCD Digital Display Vernier Caliper)
5	Clinical (assessment models cass)	Intermolar diameter difference	Intermolar diameter difference (6 diameter)	Millimeters	Digital Caliper (0–150mm/0–6″ Large LCD Digital Display Vernier Caliper)
6	Clinical	Palate rotation of upper molars	Molars rotation	Binary (absence = 0; presence = 1)	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
7	Clinical	Canine class malocclusion right	Canine class malocclusion right	first = 0; second = 1; third = 2	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
8	Clinical	Canine class malocclusion left	Canine class malocclusion left	first = 0; second = 1; third = 2	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
9	Clinical	Molar class malocclusion right	Molar class malocclusion right	first = 0; second = 1; third = 2	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
10	Clinical	Molar class malocclusion left	Molar class malocclusion left	first = 0; second = 1; third = 2	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
11	Radiographic, *skeletal*	Sella–nasion–A point angle	SNA	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
12	Radiographic, skeletal	Sella–nasion–B point angle	SNPg	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany and the computer)
13	Radiographic, skeletal	Angle formed by the NA floor with NPG plane. Sagittal intermaxillary relationship	ANPg	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
14	Radiographic, skeletal	Angle formed by sellar plane with the palatal plane ANS-PNS	SN ˆAns-Pns	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
15	Radiographic, skeletal	Angle formed by the saddle plan SN with mandibular plan Go- Gn	SNˆGo-Gn	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
16	Radiographic, skeletal	Angle formed by the palatine plan ANS – PNS with mandibular plan Go- Gn	Ans-PnsfnˆˆGo-Gn	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
17	Radiographic, skeletal	Angle formed by the palatine plan ANS – PNS with the incisor upper axis Is	Ans-PnsˆIs	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
18	Radiographic, skeletal	Angle formed by the mandibular plan Go- Gn with the lower incisor axis Ii.	Go-GnˆLi	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
19	Radiographic, skeletal	Distance between the dental plan Apg and lower incisor edge B1	Apg-B1	Millimeters	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)

(*continued on next page*)

Table A1 (*continued*)

ID	Feature category	Feature full name	Feature name/ acronym	Feature value	How it is done (tools and material employed)
20	Radiographic, skeletal	Lower Incisor position	LIP	Millimeters	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
21	Radiographic, skeletal	Upper Incisor position	UIP	Millimeters	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
22	Clinical	Overjet	OVJ	Millimeters	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
23	Clinical	Overbite	OVB	Millimeters	Intraoral photographs (Nikon 1 J5 camera), clinical evaluation and evaluation of the upper and lower dental casts (use of mechanical chair, Halogen light reflector)
24	Radiographic, skeletal	Interincisal Angle Isˆli	Isˆli	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
25	Radiographic, skeletal	Protrusion lower lip	PLI	Millimeters	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
26	Radiographic, skeletal	Co-Go-Me angle	Co-Go-Me	Degree	lateral cephalometric radiographs and cephalometric analysis (Use of the imaging system Orthopos XG 3Dready Ceph with a CCD line sensor (Sirona Dental Systems, Bensheim, Germany) and the computer)
27	Clinical (assessment models cass)	Anterior bolton index	Anterior Bolton index	(normal = 0; increased = 1; decreased = 2)	Evaluation of the upper and lower dental casts
28	Clinical	Kind of gingival	Gingival tipology (Geng Tip)	Binary (thick = 0; thin = 1)	clinical evaluation (use of mechanical chair, Halogen light reflector)
29	Clinical	Gingival recessions	Gingival recessions' presence (Rec)	Binary (absence = 0; presence = 1)	clinical evaluation (use of mechanical chair, Halogen light reflector)
30	Clinical	Labial incompetence	Labial incompetence (Lab incomp)	(absence = 0; mild = 1; severe = 2)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
31	Clinical	Aesthetic line	Aesthetic line	(orthognathic = 0; retrusive = 1; protruded = 2)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
32	Clinical	Smile teeth exposure	Smile teeth exposure	(Good = 0; scarce = 1)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
33	Clinical	Coincidence of the facial midline with the dental midline	Midline coincidence	(Coinciding = 0; not coinciding = 1)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
34	Clinical	Angle formed by a tangent line to the point subnasal and one tangent to labial filter	Nasolabial angle	(normal = 0; closed = 1; open:2)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
35	Clinical	Trend of the profile that can be normal, concave or convex	Facial profile	(normal = 0, concave = 1; convex = 2)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
36	Clinical	Distance between the chin and neck	Chin-neck distance	(Normal = 0; increased = 1; decreased = 2)	clinical evaluation and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
37	Clinical	Fixed, functional and fixed lingual	Treatment of orthodontic with extractions	(no = 0; yes D = 1, yes P = 2)	clinical evaluation, intraoral and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector)
38	Clinical (class label)	Teeth extracted	Teeth extracted	yes = 1, no = 0	clinical evaluation, intraoral and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector
39	Clinical	Fixed, functional, clear aligner and fixed lingual	Kind of treatment of orthodontic	(fixed = 0; aligner = 1; functional = 2; fixed lingual = 3)	clinical evaluation, intraoral and extraoral photographs (Nikon 1 J5 camera, use of mechanical chair, Halogen light reflector

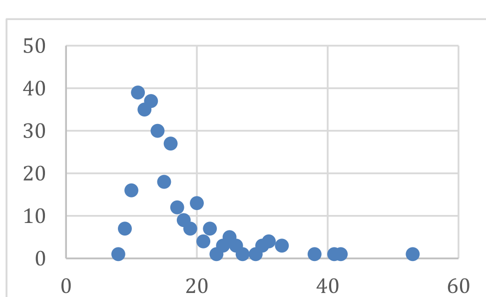

Fig. A2. Patient's age distribution. *Ages* are on the x-axis while their *frequency* is on the y-axis.

Appendix B

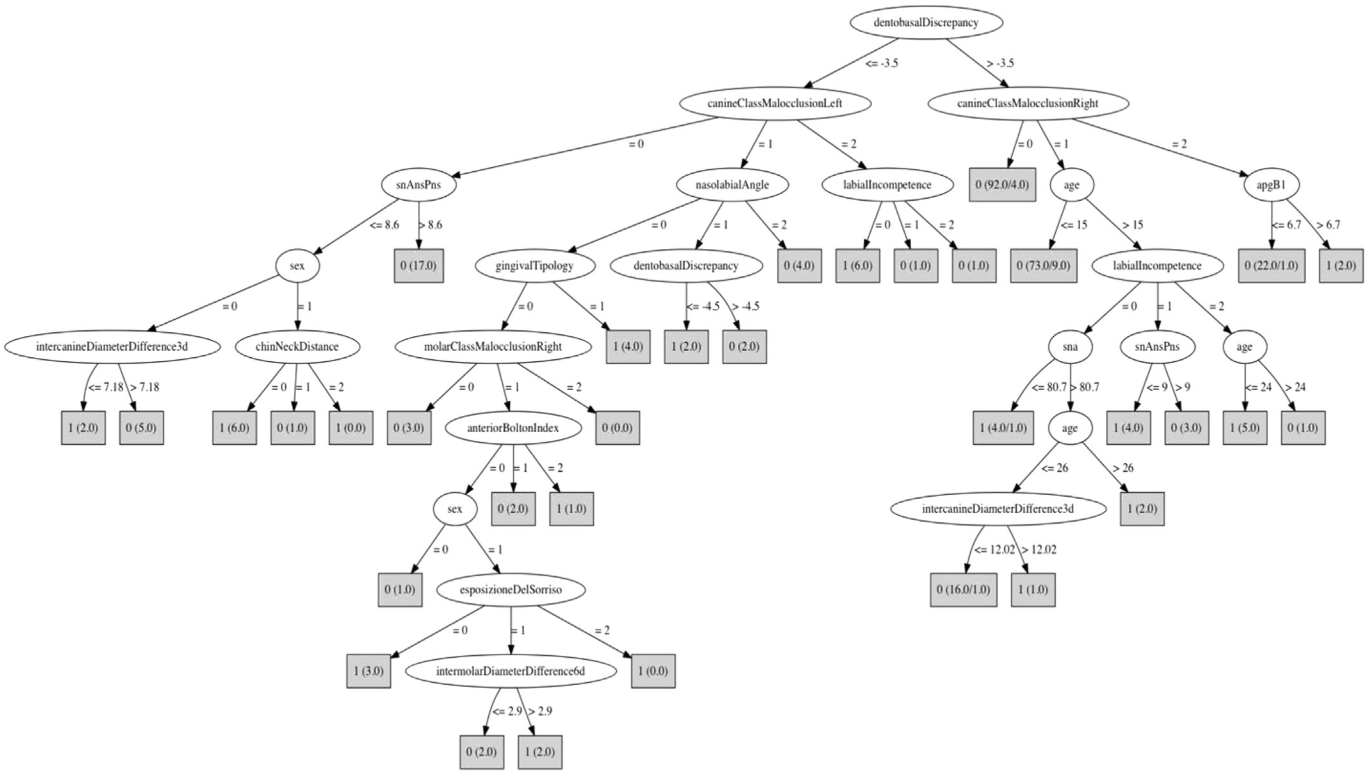

Fig. B1. *Decision tree* of the first experimental setting.

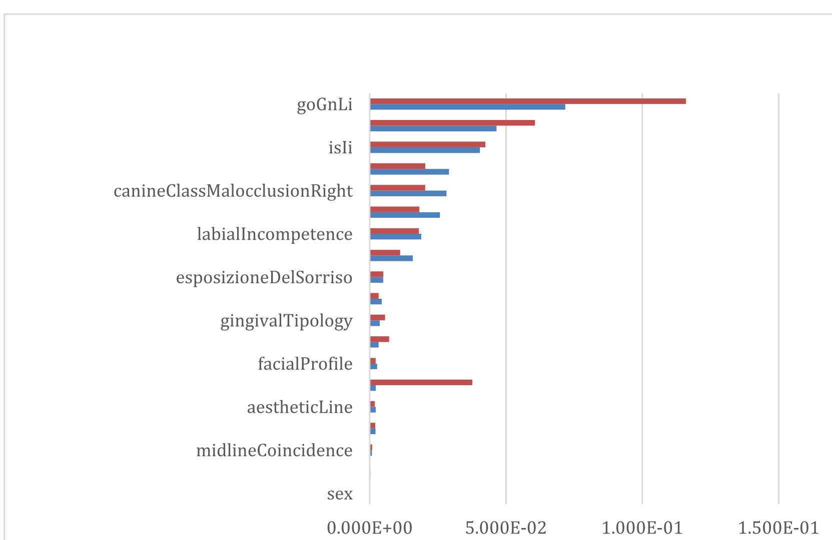

Fig. B2. Attribute ranking comparison for the first experimental setting.

Appendix C

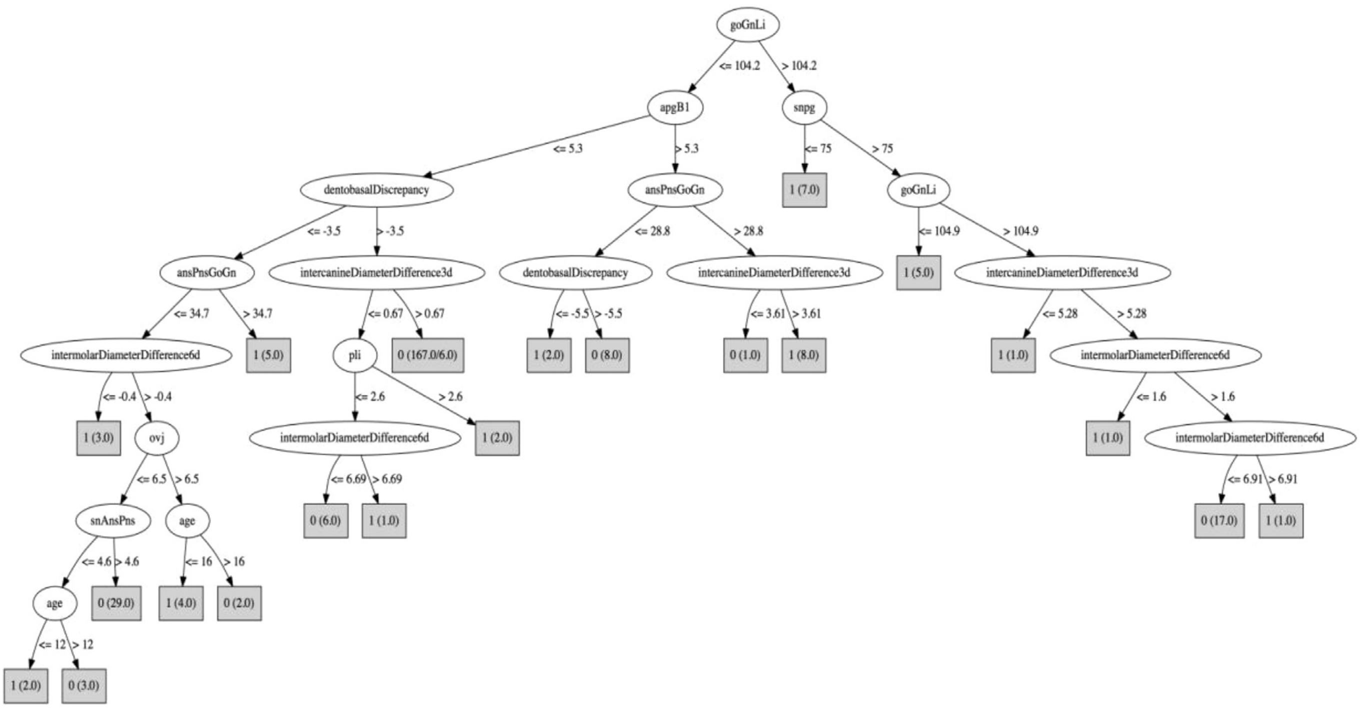

Fig. C1. *Decision tree* of the second experimental setting.

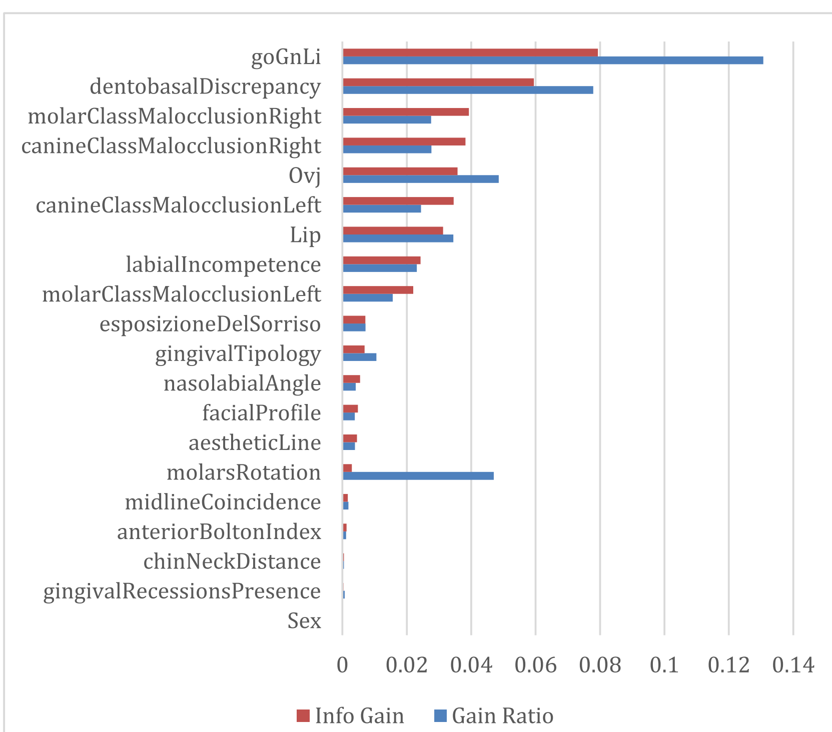

Fig. C2. Attribute ranking for the second experimental setting.

Appendix D

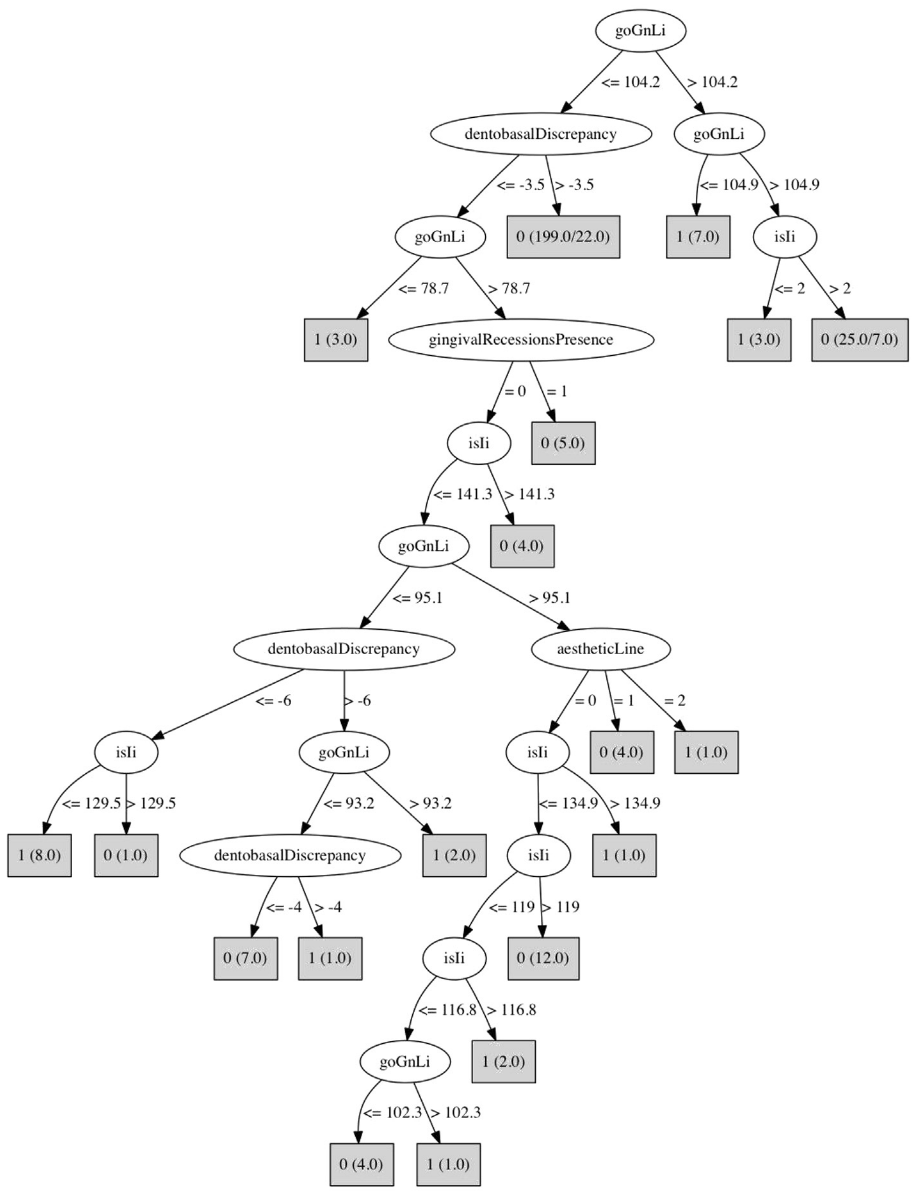

Fig. D1. *Decision tree* of the third experimental setting.

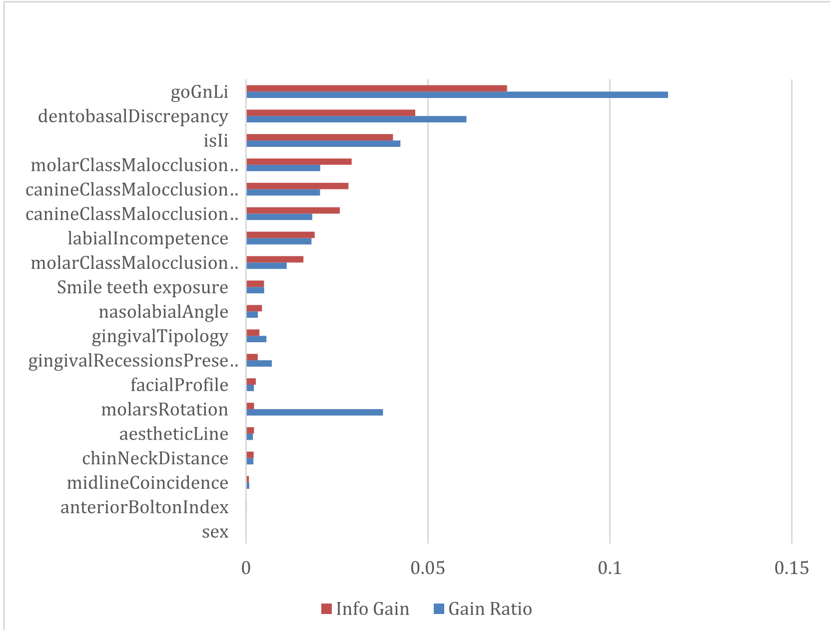

Fig. D2. Attribute ranking comparison for the third experimental setting.

References

Anderson, D.R., Sweeney, D.J., Williams, T.A., Camm, J.D., Cochran, J.J., 2015. An introduction to management science: quantitative approaches to decision making. Cengage Learn.

Bodemer, N., Hanoch, Y., Katsikopoulos, K.V., 2015. Heuristics: foundations for a novel approach to medical decision making. Intern. Emerg. Med. 10 (2), 195–203.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Croskerry, P., 2015. Clinical decision making. In: pediatric and congenital cardiac care. Springer, London, pp. 397–409.

Faezipour, M., Ferreira, S., 2013. A system dynamics perspective of patient satisfaction in healthcare. Procedia Comput. Sci. 16, 148–156.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 (8), 861–874.

Figueroa, R.L., Zeng-Treitler, Q., Kandula, S., Ngo, L.H., 2012. Predicting sample size required for classification performance. BMC Med. Inf. Decis. Mak. 12 (1), 8.

Hammond, J., Keeney, R., Raiffa, H., 2015. Smart Choices: a Practical Guide to Making Better Decisions. Harvard Business Review Press.

Heinrichs, H., Laws, N., 2014. "Sustainability state" in the Making? Institutionalization of sustainability in German federal policy making. Sustainability 6, 2623–2641.

Hu, Z., Yin, X., Liao, J., Zhou, C., Yang, Z., Zou, S., 2015. The effect of teeth extraction for orthodontic treatment on the upper airway: a systematic review. Sleep Breath. 19 (2), 441–451.

Jameton, A., McGuire, C., 2002. Toward sustainable health-care services: principles, challenges, and a process. Int. J. Sustain. High. Educ. 3 (2), 113–127.

Kravitz, N.D., Bowman, S.J., 2016. A paradigm shift in orthodontic marketing. In: Seminars in Orthodontics. WB Saunders.

Lin, F., Ren, M., Yao, L., He, Y., Guo, J., Ye, Q., 2016. Psychosocial impact of dental esthetics regulates motivation to seek orthodontic treatment. Am. J. Orthod. Dentofac. Orthop. 150 (3), 476–482.

Lozano, R., Lukman, R., Lozano, F.J., Huisingh, D., Lambrechts, W., 2013. Declarations for sustainability in higher education: becoming better leaders, through addressing the university system. J. Clean. Prod. 48, 10–19.

Martina, R., Teti, R., Addona, D.D., Iodice, G., 2004. Neural Network Based System for Decision Making Support in Orthodontic Extractions. I* PROMS.

Masías, V.H., Krause, M., Valdés, N., Pérez, J.C., Laengle, S., 2015. Using decision trees to characterize verbal communication during change and stuck episodes in the therapeutic process. Front. Psychol. 6, 379.

Matos, S., Silvestre, B.S., 2013. Managing stakeholder relations when developing sustainable business models: the case of the Brazilian energy sector. J. Clean. Prod. 45, 61–73.

Mitchell, T.M., 1997. Machine Learning, 1 ed. McGraw-Hill, Inc., New York, NY, USA.

Mulimani, P., 2017. Green dentistry: the art and science of sustainable practice. Br. Dent. J. 222, 954–961.

Musen, M.A., Middleton, B., Greenes, R.A., 2014. Clinical decision-support systems. In: Biomedical Informatics. Springer, London, UK, pp. 643–674.

Nguyen, T., Proffit, W., 2016. The Decision-making Process in Orthodontics. Current Principles and Techniques 208, Orthodontics.

Nouri, M., Asefi, S., Baghban, A.A., Ahmadvand, M., Shamsa, M., 2016. Objective vs subjective analyses of arch form and preformed archwire selection. Am. J. Orthod. Dentofac. Orthop. 149, 543–554.

Novo-Corti, I., González-Laxe, F., Pociovalisteanu, D.M., 2015. The economic analysis of maritime catastrophes in sensitive areas: the assessment and calculation of damages in the environment and population's way of life. J. Clean. Prod. 107, 267–278.

Pithon, M.M., de Faria, L.C.M., Tanaka, O.M., Ruellas, A.C., de, O., Primo, L.S., de, S.G., 2017. Sustainability in Orthodontics: what can we do to save our planet? Dent. Press J. Orthod. 22 (4), 113–117.

Quinlan, R.J., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Ribarevski, R., Vig, P., Vig, K.D., Weyant, R., O'Brien, K., 1996. Consistency of orthodontic extraction decisions. Eur. J. Orthod. 18 (1), 77–80.

Sammut, C., Webb, G.I., 2011. Encyclopedia of Machine Learning. Springer Science & Business Media.

Silvestre, B.S., 2015. A hard nut to crack! Implementing supply chain sustainability in an emerging economy. J. Clean. Prod. 96, 171–181.

Stiggelbout, A.M., Van der Weijden, T., De Wit, M.P.T., Frosch, D., Legare, F., Montori, V.M., Elwyn, G., 2012. Shared decision making: really putting patients at the centre of healthcare. BMJ Br. Med. J. (Online) 344.

Straus, S.E., Sackett, D.L., 1998. Getting research findings into practice: using research findings in clinical practice. BMJ 317 (7154), 339.

Turpin, D.L., Huang, G., 2016. The role of evidence in orthodontics. Orthod. Curr. Princ. Tech. 200.

Wals, A.E., 2014. Sustainability in higher education in the context of the UN DESD: a review of learning and institutionalization processes. J. Clean. Prod. 62, 8–15.

Weiss, G.M., 2004. Mining with rarity: a unifying framework. ACM SIGKDD Explor. Newsl. 6 (1), 7–19.

Witten, I.H., Frank, E., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Xie, X., Wang, L., Wang, A., 2010. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. Angle Orthod. 80 (2), 262–266.