

Perspective

# The Validity of Machine Learning Procedures in Orthodontics: What Is Still Missing?

Pietro Auconi <sup>1</sup>, Tommaso Gili <sup>2,3,\*</sup>, Silvia Capuani <sup>3</sup> , Matteo Saccucci <sup>4</sup> , Guido Caldarelli <sup>3,5,6</sup>, Antonella Polimeni <sup>4</sup>  and Gabriele Di Carlo <sup>4</sup> 

<sup>1</sup> Private Practice of Orthodontics, 00012 Rome, Italy; pietroauc@gmail.com

<sup>2</sup> Networks Unit, IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy

<sup>3</sup> ISC CNR, Department of Physics, University of Rome "Sapienza", P.le Aldo Moro 5, 00185 Rome, Italy; silvia.capuani@isc.cnr.it (S.C.); guido.caldarelli@unive.it (G.C.)

<sup>4</sup> Department of Oral and Maxillo-Facial Sciences, Sapienza University of Rome, Viale Regina Elena 287a, 00161 Rome, Italy; matteo.saccucci@uniroma1.it (M.S.); antonella.polimeni@uniroma1.it (A.P.); gabriele.dicarlo@uniroma1.it (G.D.C.)

<sup>5</sup> Department of Molecular Sciences and Nanosystems, Ca' Foscari University of Venice, Via Torino 155, Venezia Mestre, 30172 Venice, Italy

<sup>6</sup> ECLT, Ca' Bottacin, Dorsoduro 3246, 30123 Venice, Italy

\* Correspondence: tommaso.gili@imtlucca.it

**Abstract:** Artificial intelligence (AI) models and procedures hold remarkable predictive efficiency in the medical domain through their ability to discover hidden, non-obvious clinical patterns in data. However, due to the sparsity, noise, and time-dependency of medical data, AI procedures are raising unprecedented issues related to the mismatch between doctors' mental reasoning and the statistical answers provided by algorithms. Electronic systems can reproduce or even amplify noise hidden in the data, especially when the diagnosis of the subjects in the training data set is inaccurate or incomplete. In this paper we describe the conditions that need to be met for AI instruments to be truly useful in the orthodontic domain. We report some examples of computational procedures that are capable of extracting orthodontic knowledge through ever deeper patient representation. To have confidence in these procedures, orthodontic practitioners should recognize the benefits, shortcomings, and unintended consequences of AI models, as algorithms that learn from human decisions likewise learn mistakes and biases.

**Keywords:** machine learning; artificial intelligence; orthodontics; complexity; prognosis optimization



**Citation:** Auconi, P.; Gili, T.; Capuani, S.; Saccucci, M.; Caldarelli, G.; Polimeni, A.; Di Carlo, G. The Validity of Machine Learning Procedures in Orthodontics: What Is Still Missing? *J. Pers. Med.* **2022**, *12*, 957. <https://doi.org/10.3390/jpm12060957>

Academic Editors: Qiang Zhang and Yifeng Zeng

Received: 21 April 2022

Accepted: 5 June 2022

Published: 11 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

For thirty years now, AI procedures have proven to be highly effective tools when implemented correctly, allowing one to perceive subtle information and ultimately convert information into actions, from speech recognition to natural language processing, spam filters, fraud detection, and many other applications [1–6]. Machine learning (ML), a subcategory of AI, is the method of creating models that perform a specific task without the need to be explicitly programmed by a human for the discovery of intricate correlations within large masses of data [7–10]. In biomedicine, ML tools have been widely applied in the handling of large numbers of patient micro-variables and in predicting the future outcomes of diseases based on previous data collected regarding similar diseases [11–17]. Despite its success, ML is a still-emerging technique in medicine, even moreso in orthodontics, and many opportunities remain unexplored.

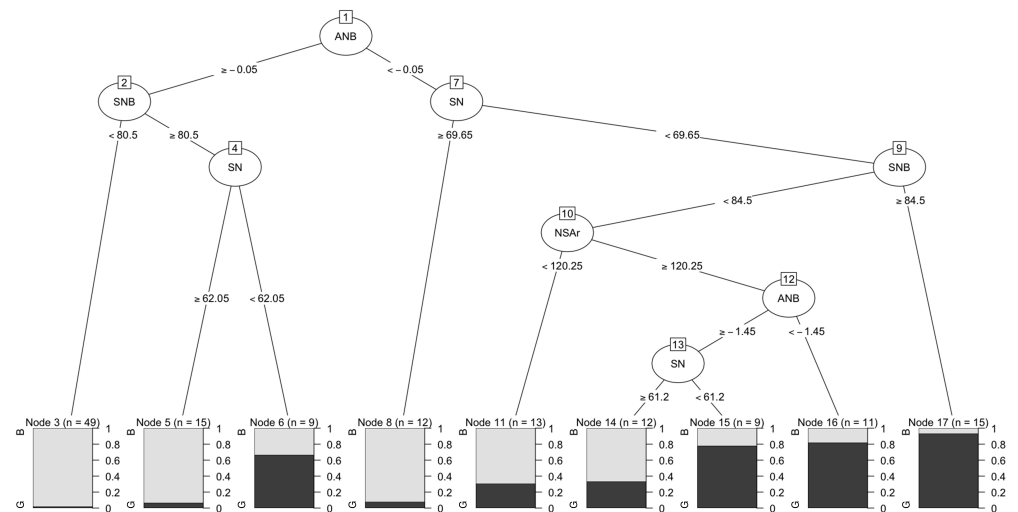
Physicians seem to overlook the fact that machine judgment is, on average, at least as reliable as expert judgments and, in many circumstances, even better [6,9]. In recent years, an excessive tendency to rely on ML systems ("over-reliance"), overdependence, "deskilling", and even potential desensitization to patient problems have been the most common criticisms related to the use of these models in medicine [18–21]. Concerns have arisen

regarding the possibility of incorporating “dirty” data and spurious correlations based on uncertain clinical interpretations [22]. Some practitioners believe that using computational systems could lead to the establishment of a new, problematic medical empiricism, based not on concrete facts and data relating to patients but rather on interconnections of data, meaning that the most straightforward clinical situations can become complicated and confusing [23]. Is there any truth to these claims? If it is true that automation will not turn us into robots (just as industrialization did not turn us into machines), is it possible that these computational procedures could increase rather than decrease the risk of statistical-clinical misunderstandings? In this paper we discuss the persisting unresolved issues related to ML computational predictive tools applied to orthodontics, which still limit the automatic extraction of valuable, clinically actionable orthodontic knowledge.

## 2. Challenging Interface between Machine Learning Models and Orthodontic Features

Many reports have highlighted the usefulness and potential of predictive electronic systems in clinical and research orthodontics [23–27]. The potential identified in these reports is similar to what has been hypothesized in medical practice—the records of the best clinical decisions made by thousands of professionals must be exploited to optimize patient care [28–30]. Although ordinary medical diagnostic approaches are based on the slow, careful recruitment of clinical and laboratory data, on subjects including the causes and effects of clinical phenomena, the significance of symptoms, and so on, the most sophisticated predictive ML implementations learn and store information at great speed, solving complex problems by repeatedly re-examining the data and layering simple concepts onto more complex ones. Translating from daily orthodontics to ML models, an orthodontist might (i) identify hidden craniofacial trends in large datasets of growing patients, (ii) leverage trends to make growth outcome predictions, (iii) compute the probability for each possible growth and treatment outcome, or (iv) clarify the effects of the co-occurrence of skeletal defects and the renormalization phenomena on growth and treatment outcomes. [28–34].

The fundamental requirement of predictive analytics procedures applied to orthodontics concerns the availability of accurate clinical data, with which the machine can gain “domain experience” [35,36]. Computer scientists use what they know how to do (algorithms) to data, the peculiarities of which they do not always understand. ML developers assume that the patient dataset used to train models is uniformly and fully representative of the target patient population. However, in medicine, not all subjects are equal. Some patients give more representative information about their clinical condition than others. Moreover, the reference data of some patients may not be 100% accurate. Thus, the process of weighing higher-quality information against other information is debatable and subjective, so the data quality dimension in orthodontic records remains a cognitively elusive concept [12,16,17]. There are differences between actual craniofacial morphology assessments, as experienced by orthodontists, and their codified representation in a numerical form, which is the case for the data input for any ML algorithm. Computational machines teach orthodontists to map medical phenomena into numerical structures in order to quantify them. Based on associations, algorithms can exploit features that the orthodontist may consider irrelevant to the problem. It is not essential to understand in depth the functioning of various patient characteristics in order to extract answers. It is necessary to engage with constant change, randomness, and noise in the data relating to orthodontic treatment and look for regularities within the data, rather than for clinical-logical hypotheses. In other words, one must allow the numbers to speak for themselves (Figure 1).



**Figure 1.** Classification trees are predictive models that can be used to decompose a problem into increasingly simple subcomponents. A tree is composed of branching processes that emerge from a series of binary selections (which are set to be larger or smaller than a reference). Tree algorithms learn through repeated exposure to clinical cases (“examples”). For classification purposes, a tree is built by repeatedly dividing the observations (e.g., cephalometric features as independent input features) into subsets that are as homogeneous as possible in relation to the dependent variable (the label). For learning to occur, data used for training must be labeled. In the example provided, the time progression of cephalometric data from 80 class III male and female growing subjects (aged from 7 to 14 years) was associated with a label indicating good (improving) or bad (worsening) craniofacial growth. The initial ANB angle reference (-0.05 degrees) was chosen to start the branching process, including other cephalometric characteristics (each associated with a specific reference). In the end, it was possible to establish whether the whole configuration was associated with bad (B) or good (G) growth. This simple procedure may constitute a prognostic aid for the orthodontic operator in communicating risk to parents. The symbolic learning related to classification trees is probably the most expressive procedure for medical data analysis when interpretability is desired. Trees were produced using the R package “tree” v1.0-37. ANB angle (degrees): measure of the relative position of the maxilla to the mandible; SNB angle (degrees): measure of the angle between the sella/nasion plane and nasion/B plane; NSAr angle (degree): measure of the angle between the anterior and posterior cranial base; SN (mm): antero-posterior length of the cranial base. Patient data were kindly offered by professors Lorenzo Franchi and James A. McNamara Jr.

Concerning the use of ML models, there are problems related to the the steps necessary to move from the observed patient data to the statements concerning future patients who have never been seen before. Machines can capture hierarchical regularities and dependencies in the data to learn complex correlations between input and output features without any inherent representation of causality [10,13,21,22]. ML procedures do not require theoretical bases. Unfortunately, predictions based on data rely on mere correlations without theories and models. A correlation quantifies the statistical relationship between the values of two parameters without clarifying the inner mechanisms of a system. Probabilities based on scarce data are not reliable, but it is not just a question of quantity. The primary source of data for the training of ML orthodontic models is data produced during growth and/or during treatment. The most difficult factor to respond to is the possible shifts between time-series data that may induce temporal drifts, which can cause algorithms to become progressively inaccurate. Craniofacial data continually evolve throughout growth, so the future does not always look like the past. Given these premises, one may suspect that the procedural ML logic applied to orthodontics may pose more than one interpretive problem.

However, sufficient evidence has accumulated that ML tools, when applied to orthodontics, perform well [22–26].

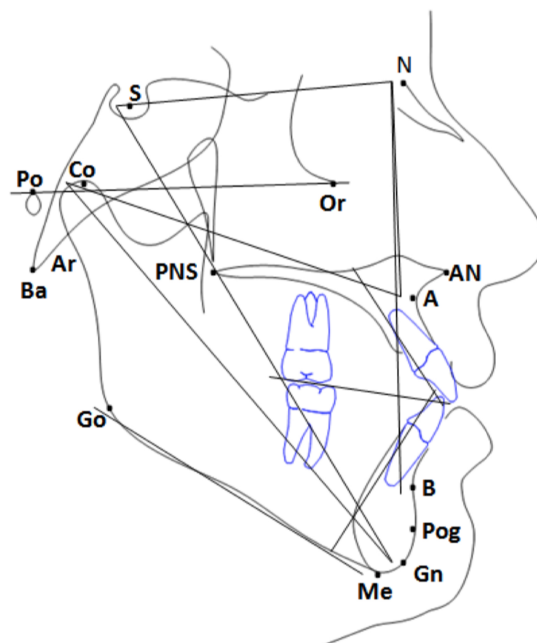
When numbers are the only object of interest in diagnostic and prognostic processes, some dehumanization occurs. Consequently, a certain amount of ineradicably intrinsic potential distortion in the interpretation of orthodontic conditions may arise. However, humans are fortunate to possess a fundamental property that the most up-to-date ML systems lack: common sense. Humans can infer the reasons behind processes, identifying abstract similarities and analogies based on only a few observed patients. Machines have facilitated disparate orthodontic clinical situations: diagnostic assessments; prognostic predictions; and the identification of salient features in growing patients at risk of skeletal imbalance, poor response to treatment, maxillofacial tumours, cysts, periapical abscesses, etc. [27–30]. The accurate localization of cephalometric landmarks using ML tools has led to a mitigation of the problem of interpersonal variation in landmark tracing and related errors in diagnosis and treatment planning [24]. The application of photography-based systems to assess jaw disharmonies (responsible for masticatory dysfunctions and apnea syndrome) and to establish the need for extractions in cases of tooth crowding and protrusion, are additional crucial steps for the successful application of ML-supported decision procedures [25,26] (see Appendix A).

### 3. How Can Orthodontic Input Be Incorporated into the Machine Learning Process?

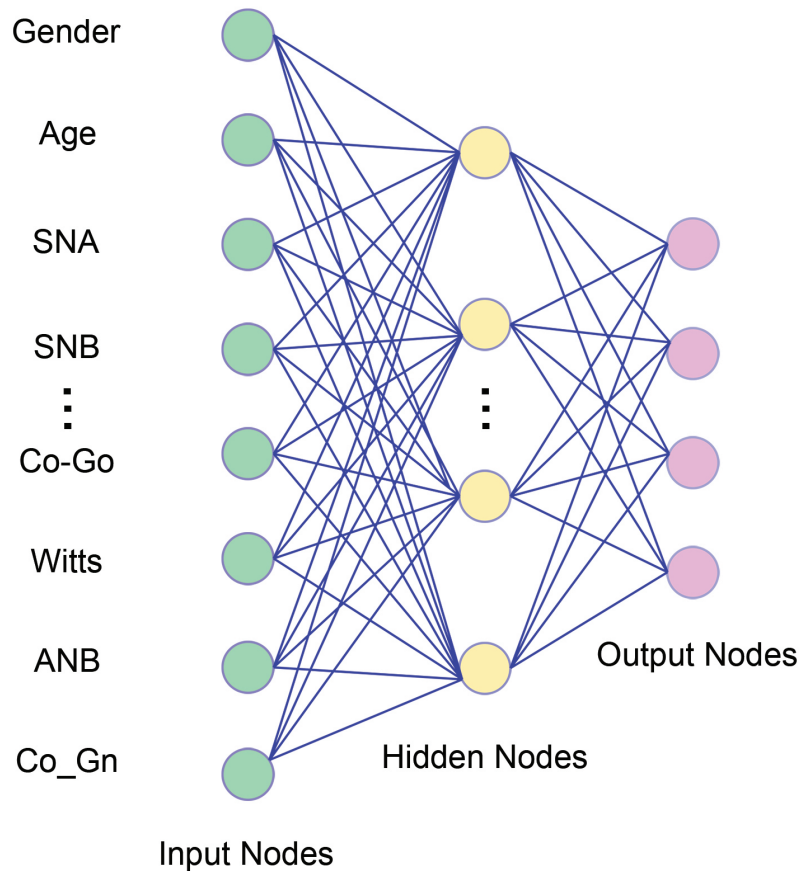
During the growth process, the clinical and cephalometric data used to feed ML machines have inherent randomness related to massively parallel processes of skeletal plasticity, which propagate through algorithms with an unavoidable degree of inaccuracy. The stochastic processes that cause a developmental trait to deviate from its expected path [20], also known as developmental noise, are an inherent part of craniofacial development and remodelling. Significant variations at the organ and whole organism level are related to the stochasticity of random intermolecular collisions, gene fluctuations, signal transduction factors, chromatin structure, DNA methylation state, morphogenetic cytoskeleton dynamics, bone translations, and other factors [20,21]. The complex pathobiology of craniofacial growth recalls a well-known saying among data scientists: all data is dirty. Nevertheless, the hypothesis offered by electronic systems is that the combination of multiple subtle aspects and a sequence of non-linear data transformations can be performed to extract both clinical and subclinical patient nuances (Figure 2), covertly containing the answer to a given problem.

In the unfolding of clinical reasoning, physicians make diagnostic errors 5%–15% of the time, depending on their speciality [16]. Two to four pieces of clinical information are sufficient to generate diagnostic hypotheses through intuition. The errors are related to the fast closure of the diagnostic process, as well as the tendency not to consider alternative views to the first diagnosis (“anchoring bias”), the tendency to consider diagnoses that are easy to remember (“availability bias”), and the tendency to include only confirmatory data for the initial diagnosis while ignoring contradictory data (“confirmatory bias”) [13,17]. Conversely, errors for machines mainly occur during the learning step. Since machines do not have the capability for intuition, the most common cause of errors lies in the poor quality of training data, such as irrelevant features, spurious associations, false assumptions, inappropriate patient attributions, and indications that are unable to represent the patient’s clinical background [9–12]. Computational models define their reality and use it to justify their results and make predictions. However, living organisms cannot be reduced to a set of mathematical equations suitable for describing an elementary mechanism; the internal parts are not endowed with the statistical homogeneity that would allow the application of probability theory. Craniofacial imbalance constitutes a repository of physical order in which a large amount of information is concentrated. Patients’ unequal developmental probabilities are due to morphological constraints, competition/collaboration strategies of skeletal elements, emergent phenomena, bone translations, and more. Patients with severe facial imbalance escape dento-alveolar renormalization systems since they tend to maintain

disharmony over time (Figure 3). Some developmental properties directionally constrain the possible path of evolution, defining the limits of the possible craniofacial variations associated with that specific initial morphology (“canalization”) [13]. Intuitively, the numerical transposition of these concepts is somewhat problematic. Algorithmic decisions are expressed in the form of rigid, not-fuzzy binary classifications (spam-not-spam, dog-cat, etc.). Making prognostic clinical predictions means identifying the presence of unfavorable factors when they have not yet occurred. To obtain a satisfying computer-assisted predictive ability, the operator must provide the machine with a series of expressive examples of the condition to be detected, i.e., examples of patients with signs and symptoms typical of the disease, which can easily be differentiated from healthy, symptom-free patients [35–38]. In the orthodontic scenario, in the same patient, shaded morphological/radiographic features of malocclusion may coexist with typical craniofacial characteristics or even with signs of a different malocclusion. There is no such vagueness in mathematical language. In mathematical language, everything is precise. Machines tend to complete information when only part of the system is known. Each orthodontic patient contains a different amount of hidden data and latent variables not expressed in numerical format (the “hidden half”) [20]. All of these affect the outcome, so two similar patients can carry a very different facial growth potential and potential responses to treatment related to different inherent amounts of developmental noise. Strains related to the imbalance between ideal prognostic models and the everyday fuzzy orthodontic reality are called “misdiagnoses” and “wrong prognoses” by orthodontists and “residuals” by statisticians. (see Appendix B).



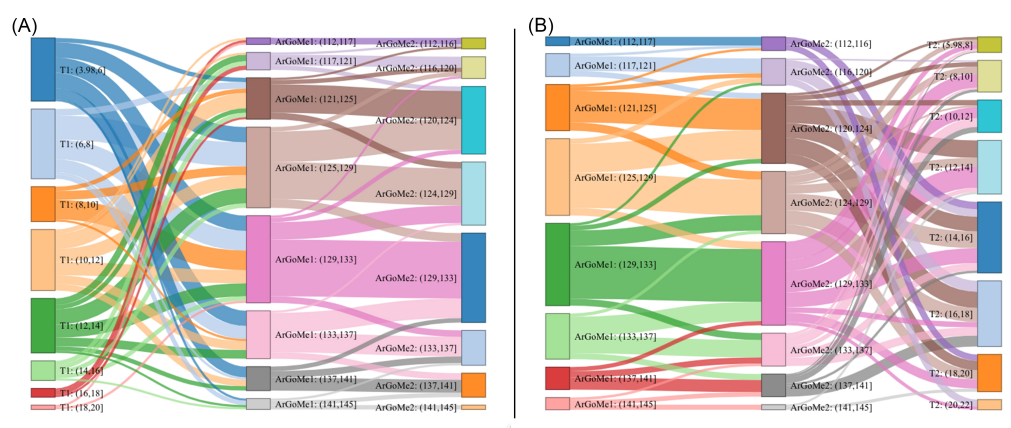
**Figure 2.** Cephalometric angular and linear measures.



**Figure 3.** A simplified Neural Network. Artificial neural network (ANN) models can “learn” from the data without any pre-specified rules and can focus mathematically on predictive performance. ANNs take the raw data at the lower (input) layer and transform them into an increasingly “abstract” representation of the characteristics. ANNs are flexible and versatile tools. A few assumptions are required about the normal distribution of errors, correlations among variables, and linear relationships among variables. They are highly applicable for any real-world situation but require many attributes and observations. The difficulties in ANN research applied to orthodontics come in many different forms. The most important contribution is the lack of a uniform feature standards in building ANN models. The second primary reason is that ANNs have fewer assumptions and many more options in the modelling process, which opens up several possibilities for their inappropriate use and applications. Deep learning is a type of ANN procedure carrying multiple Although node layers. Each layer learns the representation of data by abstracting data in many ways. While traditional statistical techniques require transforming raw data (feature engineering) to represent the problem and make predictions, deep learning algorithms achieve this automatically, using more and more abstract levels of representation, encapsulating highly complicated functions in the process.

#### 4. Tell Me What You Have Understood about This Patient

Practitioners generally trust their subjective intuition more than the answer provided by an algorithm [2,3,39,40]. Humans are poor at making probabilistic decisions based on partial information and cannot even precisely calculate how data interfere with each other [31,41]. As already mentioned, in a patient dataset, some components (for instance, dentoalveolar adaptive remodelling) can remain below the threshold of perception of ML tools [42–55]. Some features may be irrelevant, missing, or redundant. The most up-to-date deep artificial neural networks do not require any additional pre-processing; they automatically cut out uninteresting correlations between parameters to build up a meaningful subset of data (Figure 4).



**Figure 4.** Trajectories of ArGoMe angle values during the growth process in 140 patients with Class III malocclusion, divided into eight classes of age. ArGoMe (Gonial angle) is the angle between the corpus and the ramus of the mandible. The Sankey diagram is usually used to indicate a data transfer in a process. In this data visualization, the width of the arrows is proportional to the number of feature flows. Sankey diagrams (A) and (B) draw attention to the transfer of values of the ArGoMe angle between two temporal acquisitions, T1 and T2. Plot (A) shows how the eight classes of age are distributed across the ArGoMe values at T1 and their evolution towards T2; plot (B) reveals how the ArGoMe values at T2 are distributed across the classes of age at T2. The Sankey diagram was obtained in ggplot2. Patient data were kindly offered by professors Lorenzo Franchi and James McNamara Jr.

This procedure allows for the comparison of the results expected by an expert, based on experience, and what is discovered by means of computational rules. One of the fascinating elements of ML algorithms lies in their ability to attribute the degree of reliability of each prediction to a self-validation process [1–3]. The ability of machines to give proper weight to the various patient factors involved in the prediction is much larger than that of humans. Paradoxically, to optimize predictions and to generalize to as many patients as possible, the software logic requires individual patient specificities to be flattened out (the “regularization” procedure) [17,18,51]. A system that is too smart in the diagnostic process ends up focusing too much on individualized patient information and has difficulties in diagnosing new patients with only slightly different characteristics (the “overfitting” phenomenon) [1–3]. To facilitate the generalization of the model performance, sometimes there is a need to inflate the data set with confounding noise. There is also the possibility of implementing procedures of data augmentation, as ML systems can create additional “synthetic patients” to improve the accuracy of forecasts [20,52–55].

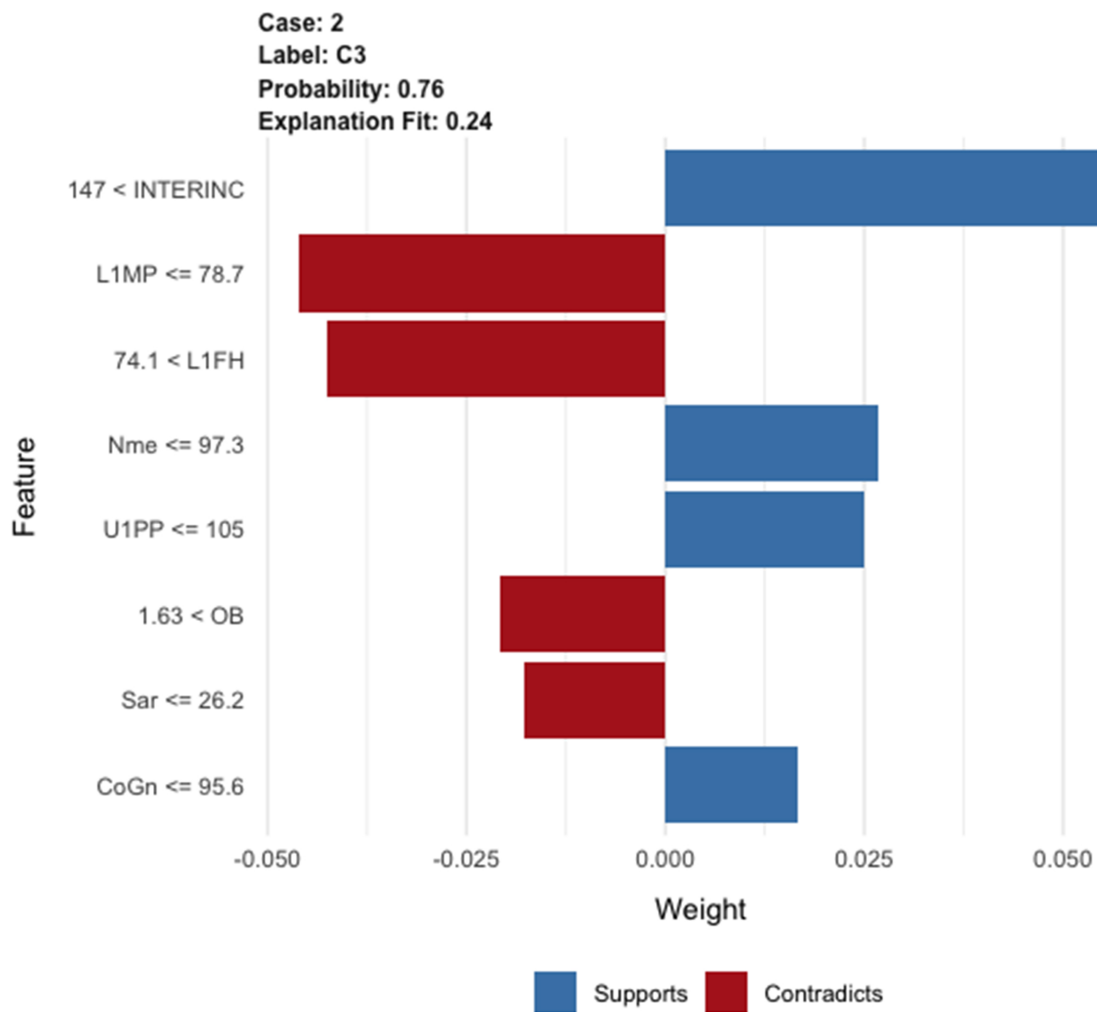
Without proper professional orthodontic supervision, the science that has allowed us to refine the patient description through feature engineering and feature selection risk may lead us to implement naïve approaches and to base decisions upon elementary clinical-technological, over-purified versions of patients. An example of the difficult balance between patient specificity, patient context, and computational answers is offered by the (sometimes too ingenuous) treatment programs underlying dental aligners. To achieve the desired outcome, a good dental alignment program, in theory, should be able to incorporate the interactions between dental movements and facial aesthetics and account for the co-occurrence of different patient characteristics, including skeletal and functional constraints, atypical swallowing, mouth breathing, and many others [24]. In the age of big data in biomedicine, it is becoming less and less possible to know in advance the direction and nature of calculations based on collective data [32–34,37–46]. Currently, fully automated methods for model selection and automatic parameter optimization are available, such as AutoML, neural architecture search, differentiable architecture search, reinforced learning, and many others [54,55]. These procedures allow the discovery of data architectures that are far more complicated than those which humans may think of trying. As there is an apparent difference between patient recognition and genuine comprehension,

the orthodontist must attempt to integrate computational responses with his cognitive cause-and-effect system carefully. Often, the advice is to broaden the patient sample to better frame the system's structure. The amount of data does not allow the consideration of fundamental questions regarding the validity of constructs such as the question of whether the patient traits are stable and comparable across patients and over time? When searching a more extensive biomedical database, it is easy to find a pattern that seems interesting, even when it is not actually relevant. Each random dataset observed over time can determine any pattern [17]. Despite the element of predictability that is missing (not expressed in the available data), relying on the algorithmic outcome prediction means trusting the ability of computational abstractions to nevertheless understand the patient by probing deeply and recursively into both visible and latent attributes. The computer-aided orthodontic operator hopes to overcome the prognostic uncertainty through repeated "deep" situational data abstractions, applying a more significant number of patients and a greater number of layers of computation.

### 5. A Matter of Trust

Human memory is an active process, based on encoding, storing, and retrieving previously acquired information [35,42]. At the chairside, orthodontists make reasoned decisions based on the logic of biomechanics and a somewhat schematic taxonomy of malocclusions. Their cognitive statistics (experience) highlight the underlying prevailing clinical trends for each patient and elements that are not very or not at all "datable", such as cultural and family aspects, compliance, and others [39]. ML statistics help orthodontists to highlight the outcome probabilities and the probabilities of escaping these trends. To disseminate the best practices and to enable researchers and practitioners to trust ML procedures, they first need to understand the bases underlying the algorithmic decisions and predictions. This would require a comprehension, at least in principle, of differences in numerical and orthodontic formalisms within the inscrutable hidden "black box" of algorithms [56]. Although the nature of ML optimization is purely mathematical, craniofacial feature optimization during growth is, above all, a matter of adaptation [57–59]. The best possible clinical-digital model may include neither the past nor the present, but only a situation calculated at every moment. When prognostic processes are conducted across both technological and morphological boundaries, new orthodontic theories could be derived through the pure power of technology [58–61]. Machines must be understandable and acceptable, even though the understandability of the algorithmic answers is often inversely proportional to the transparency and the complexity of the predictive models [46] (Figure 5). Computational models attempt to organize thoughts. Despite the necessary refinements that have been, when applied to orthodontics, these procedures have been proven to improve the professional skills of orthodontists and will soon do so even more effectively. For the benefits of a more productive man-machine operational coupling, future research should focus on a new form of digital ecology. Specifically, better interactive methods are needed for dealing with residual algorithmic standardization issues, better guidelines of algorithmic procedures, and better governance of the processes involved in the creation, validation, and updating of predictive orthodontic models.





**Figure 5.** Trust in individual predictions is crucial when the model is used for treatment decisions. Using the LIME explanation procedure [54], we can explain the predictions of any classifier or regressor by approximating it locally with an interpretable model. The figure shows the cephalometric features of one Class III male patient with very bad maxillomandibular growth. The bar chart represents the importance of the most relevant cephalometric features that supported the prediction of increasing skeletal imbalance. The green bars supported the predictions, whereas red bars contradicted them.

### 6. Conclusions

Orthodontics is characterized by prognostic uncertainty, with a strong influence of factors that are not easy to model. Therefore, reliable computerized predictive tools and procedures could be particularly welcome, even from a cautionary and medical-legal point of view. The use ML will not be able to replace orthodontists in the coming years. It will be used in cooperation with orthodontists to enhance their abilities and clinical sagacity. The significance of ML results is required to be verified repeatedly by orthodontists, patients, and computer scientists, using a stable and shared interpretative framework, in order for this technique to be more extensively applied in research and in clinical orthodontic practice.

**Author Contributions:** P.A. conceived the paper; P.A. and T.G. reviewed the bibliography; P.A., T.G., and G.D.C. wrote the paper and prepared the figures. P.A., T.G., S.C., M.S., G.C., A.P., and G.D.C. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The approval of an Ethical Committee was not sought as all data analysed were collected as part of routine orthodontic pre-treatment diagnosis.

**Informed Consent Statement:** Written informed consent was obtained from the patients' parents as part of their orthodontic treatment.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** P.A. and T.G. would like to thank M. Scazzocchio for valuable suggestions on earlier versions of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Translating from Orthodontic to ML Models

Human learning is qualitative because it is based on the interaction of meanings; each new learning experience is embedded in a process of interpretive reasoning. The learning process of a machine is instead built on the probability of the interactions of the data in the absence of prior knowledge. Orthodontic data are not stationary. The underlying processes that generate the data change over time, and so do the underlying concepts ("drifting concepts"). Invariant historical data can decrease the reliability of predictions. Because no morphological data are timeless, time trends correspond to the true stories of patients. Feature engineering (FE) is the process used to extract, aggregate, refine, and transform raw patient data into numerical features and formats that better represent the underlying problem, resulting in improved model predictive accuracy. FE aims to integrate the quantitative and qualitative aspects of a patient, to understand and structure the biological complexity through mathematical models. FE is a crucial step in the ML workflow, because the correct understanding of patient features can decrease the difficulty of modeling [18]. Enhancing the native form of the data represents the most creative phase of the predictive workflow. It is necessary to provide principles related to the domain knowledge of the operator, the intuition of the orthodontist, and skillsets regarding the advantages and disadvantages of every single computational procedure. There is a fundamental difference between representing and engineering physical features and biological features. Biological features have the capacity to evolve; this interferes with the long-term stability of the model. Perceiving these subtle aspects is the key to the successful representation of orthodontic patients. The value of a cephalometric variable can be better expressed as a distance from the population values, normalized for age and gender, rather than in its raw form. The model must therefore contain data from the context to better expose relevant aspects. Each craniofacial measurement can be related to the subject's global sagittal or vertical skeletal imbalance. A skeletal segment can be related with its morphological counterpart. It is well known that the sagittal skeletal disharmony between the maxilla and the mandible can be worsened, or mitigated, by the position of the glenoid fossa. The effect of feature co-occurrences should be included in the model as an important additional type of data. The machine cannot have any prior knowledge, and it cannot even be aware of possible concomitant dentoalveolar renormalization phenomena. The complex interplay of causes underlying atypical growth requires a different perspective on the disorders affecting the orofacial biological balance. This can be achieved by enlarging the base of information about the system's chemical, physical, and mechanical properties [62]. A regulatory or central body to prepare such data and realise an adequate database is mandatory since the provision of local data for collection within single research centers can not be sufficient.

### Appendix B. Machine Learning Programs Can Uncover Effects of Hidden Relationships between Components

To implement machine learning successfully in daily orthodontic practice we need formal rules that lead to intelligible procedures. The concept we are interested in must be

represented in the best possible way, and sometimes this is the operator's most difficult task. Concepts are best understood when they are placed in an appropriate background, so the contribution of the clinician is decisive in suggesting the degree of typicality of each patient with respect to the concept to be represented [55–57]. In order to predict the growth characteristics that lead to a malocclusion early, the learning machine must be provided with many typical examples of subjects of various ages and affected by different skeletal disharmonies. Next, time-series of features belonging to subjects with different growth trajectories are provided to the model, appropriately contrasted with features from normal subjects. When the training set changes over time, predictions tend to be ineffective. Hidden changes in context cause problems for any ML approach that assumes concept stability. Challenges involved with malocclusion mapping into ML models are as follows: (1) how to mathematize the overall skeletal imbalance; (2) how to define the background of the problem, and the specific contribution of each characteristic; (3) how to valorize the patients most expressive of the problem and how to assign feature weights; (4) how to choose different ML analytical methods designed to examine specific issues (i.e., related to growth, to habits, to the response to therapy, and so on); (5) how to identify subgroups of homogeneous patients and related risks of the occurrence and progression of malocclusion; and (6) how to capture the latent dimensions of unfavorable growth/unsuccessful treatment. The level of uncertainty that accompanies the collection of orthodontic data is usually remarkable. The programmer must include in the model the right amount of unavoidable uncertainty and bias about the data, in order to avoid excessive adherence to idealized situations. Excessive orthodontic data cleaning offers the learner an unrealistic, oversimplified representation of reality.

## References

1. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358.
2. Deo, R.C. Machine learning in medicine. *Circulation* **2015**, *132*, 1920–1930.
3. Handelman, G.S.; Kok, H.K.; Chandra, R.V.; Razavi, A.H.; Lee, M.J.; Asadi, H. eDoctor: Machine Learning and the Future of Medicine. *J. Intern. Med.* **2018**, *284*, 603–619. <https://doi.org/10.1111/joim.12822>.
4. Weinberger, D. *Everyday Chaos: Technology, Complexity, and How We're Thriving in a New World of Possibility*; Harvard Business Press: Boston, MA, USA, 2019.
5. Obermeyer, Z.; Lee, T.H. Lost in Thought—The Limits of the Human Mind and the Future of Medicine. *N. Engl. J. Med.* **2017**, *377*, 1209–1211. <https://doi.org/10.1056/nejmp1705348>.
6. Miotto, R.; Li, L.; Kidd, B.A.; Dudley, J.T. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci. Rep.* **2016**, *6*, 26094. <https://doi.org/10.1038/srep26094>.
7. Bellazzi, R.; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inform.* **2008**, *77*, 81–97. <https://doi.org/10.1016/j.ijmedinf.2006.11.006>.
8. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>.
9. Sajda, P. Machine learning for detection and diagnosis of disease. *Annu. Rev. Biomed. Eng.* **2006**, *8*, 537–565. <https://doi.org/10.1146/annurev.bioeng.8.061505.095802>.
10. Kononenko, I. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* **2001**, *23*, 89–109. [https://doi.org/10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X).
11. Freitas, A.A. Understanding the crucial role of attribute interaction in data mining. *Artif. Intell. Rev.* **2001**, *16*, 177–199. <https://doi.org/10.1023/A:1011996210207>.
12. Bzdok, D.; Altman, N.; Krzywinski, M. Points of Significance: Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. <https://doi.org/10.1038/nmeth.4642>.
13. Goldemberg, J.; Ferguson, C.; Prud'homme, A. *The World's Energy Supply: What Everyone Needs to Know*; Oxford University Press: Oxford, UK, 2015.
14. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. <https://doi.org/10.1038/nature14539>.
15. Arik, S.Ö.; Ibragimov, B.; Xing, L. Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imaging* **2017**, *4*, 014501. <https://doi.org/10.1117/1.jmi.4.1.014501>.
16. Marcus, G.; Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust*; Pantheon Books: New York, 2019.
17. Finlay, S. *Predictive Analytics, Data Mining and Big Data*; Palgrave Macmillan: New York, 2014; p. 248. <https://doi.org/10.1057/9781137379283>.
18. Zheng, A.; Casari, A. *Feature engineering for machine learning*; Number September, 2018; p. 218.
19. Kelso, J.A.S.; Engstrom, D.A. *The Complementary Nature*; MIT Press: Cambridge, MA, 2018; p.317. <https://doi.org/10.7551/mitpress/1988.00>

20. Blastland, M. *The Hidden Half: The Unseen Forces That Influence Everything*; Atlantic Books: London, 2020.
21. Pelaccia, T.; Forestier, G.; Wemmert, C. Deconstructing the diagnostic reasoning of human versus artificial intelligence. *CMAJ* **2019**, *191*, E1332–E1335. <https://doi.org/10.1503/cmaj.190506>.
22. Bichu, Y.M.; Hansa, I.; Bichu, A.Y.; Premjani, P.; Flores-Mir, C.; Vaid, N.R. Applications of artificial intelligence and machine learning in orthodontics: A scoping review. *Prog. Orthod.* **2021**, *22*, 18.
23. Nanda, S.B.; Kalha, A.S.; Jena, A.K.; Bhatia, V.; Mishra, S. Artificial neural network (ANN) modeling and analysis for the prediction of change in the lip curvature following extraction and non-extraction orthodontic treatment. *J. Dent. Spec.* **2015**, *3*, 217. <https://doi.org/10.5958/2393-9834.2015.00002.9>.
24. Asiri, S.N.; Tadlock, L.P.; Schneiderman, E.; Buschang, P.H. Applications of artificial intelligence and machine learning in orthodontics. *APOS Trends Orthod.* **2020**, *10*, 17–24. [https://doi.org/10.25259/APOS\\_117\\_2019](https://doi.org/10.25259/APOS_117_2019).
25. Li, P.; Kong, D.; Tang, T.; Su, D.; Yang, P.; Wang, H.; Zhao, Z.; Liu, Y. Orthodontic Treatment Planning based on Artificial Neural Networks. *Sci. Rep.* **2019**, *9*, 2037. <https://doi.org/10.1038/s41598-018-38439-w>.
26. Allareddy, V.; Rengasamy Venugopalan, S.; Nalliah, R.P.; Caplin, J.L.; Lee, M.K.; Allareddy, V. Orthodontics in the era of big data analytics. *Orthod. Craniofacial Res.* **2019**, *22*, 8–13. <https://doi.org/10.1111/ocr.12279>.
27. Bahaa, K.; Noor, G.; Yousif, Y. The Artificial Intelligence Approach for Diagnosis, Treatment and Modelling in Orthodontic. In *Principles in Contemporary Orthodontics*; InTech: London, 2011. <https://doi.org/10.5772/19597>.
28. Faber, J.; Faber, C.; Faber, P. Artificial intelligence in orthodontics. *APOS Trends Orthod.* **2019**, *9*, 201–205. [https://doi.org/10.25259/apos\\_123\\_2019](https://doi.org/10.25259/apos_123_2019).
29. Murata, S.; Lee, C.; Tanikawa, C.; Date, S. Towards a fully automated diagnostic system for orthodontic treatment in dentistry. In Proceedings of the 13th IEEE International Conference on eScience, eScience 2017, Auckland, New Zealand, 24–27 October 2017; pp. 1–8. <https://doi.org/10.1109/eScience.2017.12>.
30. Lux, C.J.; Stellzig, A.; Volz, D.; Jäger, W.; Richardson, A.; Komposch, G. A neural network approach to the analysis and classification of human craniofacial growth. *Growth Dev. Aging* **1998**, *62*, 95–106.
31. Deo, R.C.; Nallamothu, B.K. Learning about Machine Learning: The Promise and Pitfalls of Big Data and the Electronic Health Record. *Circ. Cardiovasc. Qual. Outcomes* **2016**, *9*, 618–620. <https://doi.org/10.1161/CIRCOUTCOMES.116.003308>.
32. Obermeyer, Z.; Emanuel, E.J. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *N. Engl. J. Med.* **2016**, *375*, 1216–1219. <https://doi.org/10.1056/nejmp1606181>.
33. Ledley, R.S.; Lusted, L.B. Reasoning foundations of medical diagnosis. *Science* **1959**, *130*, 9–21. <https://doi.org/10.1126/science.130.3366.9>.
34. Holzinger, A. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. *IEEE Intell. Inform. Bull.* **2014**, *15*, 6–14.
35. Wood, R.; Baxter, P.; Belpaeme, T. A review of long-term memory in natural and synthetic systems. *Adapt. Behav.* **2012**, *20*, 81–103. <https://doi.org/10.1177/1059712311421219>.
36. Crawford, J.; Greene, C.S. Incorporating biological structure into machine learning models in biomedicine. *Curr. Opin. Biotechnol.* **2020**, *63*, 126–134. <https://doi.org/10.1016/j.copbio.2019.12.021>.
37. Zitnik, M.; Nguyen, F.; Wang, B.; Leskovec, J.; Goldenberg, A.; Hoffman, M.M. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Inf. Fusion* **2019**, *50*, 71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>.
38. Saria, S.; Butte, A.; Sheikh, A. Better medicine through machine learning: What’s real, and what’s artificial? *PLoS Med.* **2018**, *15*, e1002721. <https://doi.org/10.1371/journal.pmed.1002721>.
39. Martínez-Abraín, A. Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. *Acta Oecologica* **2008**, *34*, 9–11. <https://doi.org/10.1016/j.actao.2008.02.004>.
40. Lovell, D.P. Biological importance and statistical significance. *J. Agric. Food Chem.* **2013**, *61*, 8340–8348. <https://doi.org/10.1021/jf401124y>.
41. Bray, D. Limits of computational biology. *Silico Biol.* **2015**, *12*, 1–7. <https://doi.org/10.3233/ISB-140461>.
42. Auconi, P.; Scazzocchio, M.; Defraia, E.; Mcnamara, J.A.; Franchi, L. Forecasting craniofacial growth in individuals with class III malocclusion by computational modelling. *Eur. J. Orthod.* **2014**, *36*, 207–216. <https://doi.org/10.1093/ejo/cjt036>.
43. Barelli, E.; Ottaviani, E.; Auconi, P.; Caldarelli, G.; Giuntini, V.; McNamara, J.A.; Franchi, L. Exploiting the interplay between cross-sectional and longitudinal data in Class III malocclusion patients. *Sci. Rep.* **2019**, *9*, 6189. <https://doi.org/10.1038/s41598-019-42384-7>.
44. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
45. Kursu, M.B.; Rudnicki, W.R. Feature selection with the boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13. <https://doi.org/10.18637/jss.v036.i11>.
46. IGuyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
47. Baumrind, S. Clinical judgment versus prediction: Towards a new paradigm for orthodontic research. In *Science and Clinical Judgment in Orthodontics*; Vig PS, R.K., Ed.; Center for Human Growth and Development, The University of Michigan, Ann Arbor, MI, 1985; pp. 149–162.

48. Auconi, P.; McNamara, J.A.; Franchi, L. Computer-aided heuristics in orthodontics. *Am. J. Orthod. Dentofac. Orthop.* **2020**, *158*, 856–867. <https://doi.org/10.1016/j.ajodo.2019.10.018>.
49. Gigerenzer, G.; Brighton, H. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Top. Cogn. Sci.* **2009**, *1*, 107–143. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>.
50. Cabitza, F.; Ciucci, D.; Rasoini, R. A giant with feet of clay: On the validity of the data that feed machine learning in medicine. In *Lecture Notes in Information Systems and Organisation*; Springer: Cham, Switzerland, 2019; Volume 28, pp. 121–136. [https://doi.org/10.1007/978-3-319-90503-7\\_10](https://doi.org/10.1007/978-3-319-90503-7_10).
51. Benítez, J.M.; Castro, J.L.; Requena, I. Are artificial neural networks black boxes? *IEEE Trans. Neural Netw.* **1997**, *8*, 1156–1164. <https://doi.org/10.1109/72.623216>.
52. Zhang, G.P. Avoiding pitfalls in neural network research. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 3–16. <https://doi.org/10.1109/TSMCC.2006.876059>.
53. Lipton, Z.C. The Mythos of Model Interpretability. *Queue* **2018**, *16*, 31–57. <https://doi.org/10.1145/3236386.3241340>.
54. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
55. Kim, B.; Khanna, R.; Koyejo, O. Examples are not enough, learn to criticize! Criticism for interpretability. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29, pp. 2288–2296.
56. Bien, J.; Tibshirani, R. Prototype selection for interpretable classification. *Ann. Appl. Stat.* **2011**, *5*, 2403–2424. <https://doi.org/10.1214/11-AOAS495>.
57. Bergadano, F.; Matwin, S.; Michalski, R.S.; Zhang, J. Learning two-tiered descriptions of flexible concepts: The POSEIDON system. *Mach. Learn.* **1992**, *8*, 5–43. <https://doi.org/10.1007/bf00994004>.
58. Vassie, K.; Morlino, G. Natural and artificial systems: Compare, model or engineer? In *Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7426, pp. 1–11. [https://doi.org/10.1007/978-3-642-33093-3\\_1](https://doi.org/10.1007/978-3-642-33093-3_1).
59. Cabitza, F.; Rasoini, R.; Gensini, G.F. Unintended consequences of machine learning in medicine. *Jama* **2017**, *318*, 517–518. <https://doi.org/10.1001/jama.2017.7797>.
60. Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Mag.* **2008**, *16*, 1–2.
61. Shortliffe, E.H.; Buchanan, B.G. A model of inexact reasoning in medicine. *Math. Biosci.* **1975**, *23*, 351–379. [https://doi.org/10.1016/0025-5564\(75\)90047-4](https://doi.org/10.1016/0025-5564(75)90047-4).
62. Di Carlo, G.; Gili, T.; Caldarelli, G.; Polimeni, A.; Cattaneo, P.M. A community detection analysis of malocclusion classes from orthodontics and upper airway data. *Orthod. Craniofacial Res.* **2021**, *24*, 172–180.