Complexity and data mining in dental research: A network medicine perspective on interceptive orthodontics

Tommaso Gili1,2 | Gabriele Di Carlo3 | Silvia Capuani2 | Pietro Auconi4 | Guido Caldarelli2,5 | Antonella Polimeni3

1Networks Unit, IMT School for Advanced Studies Lucca, Lucca, Italy
2CNR-ISC Unità Sapienza, Rome, Italy
3Department of Oral and Maxillo-Facial Sciences, Sapienza University of Rome, Rome, Italy
4Private Practice Rome, Rome, Italy
5Department of Molecular Sciences and Nanosystems, Ca’Foscari University of Venice, Venezia Mestre, Italy

Correspondence
Gabriele Di Carlo, Department of Oral and Maxillo-Facial Sciences, Sapienza University of Rome, Viale Regina Elena 287a, 00161 Rome, Italy.
Email: gabriele.dicarlo@uniroma1.it

Abstract
Procedures and models of computerized data analysis are becoming researchers’ and practitioners’ thinking partners by transforming the reasoning underlying biomedicine. Complexity theory, Network analysis and Artificial Intelligence are already approaching this discipline, intending to provide support for patient’s diagnosis, prognosis and treatments. At the same time, due to the sparsity, noisiness and time-dependency of medical data, such procedures are raising many unprecedented problems related to the mismatch between the human mind’s reasoning and the outputs of computational models. Thanks to these computational, non-anthropocentric models, a patient’s clinical situation can be elucidated in the orthodontic discipline, and the growth outcome can be approximated. However, to have confidence in these procedures, orthodontists should be warned of the related benefits and risks. Here we want to present how these innovative approaches can derive better patients’ characterization, also offering a different point of view about patient’s classification, prognosis and treatment.

KEYWORDS
artificial intelligence, big data, complexity, machine learning, musculoskeletal magnetic resonance imaging, network medicine, orthodontics

INTRODUCTION

The technological revolution we have been experiencing in the last decades impacted medical sciences in various ways. Many results from high-throughput experiments and the increased connection between data scientists have made available an unprecedented amount of insights to researchers and practitioners. Consequently, it becomes evident that many biological processes, as craniofacial growth, are prominent examples of complex systems in which the outcome depends upon the interaction of various components.1,2 Each element depends on its past properties, the properties of the other elements, and their past. This new scientific paradigm triggered the search for mathematical instruments to describe and possibly drive such complex behaviour.3 A complex system of interacting agents takes a natural mathematical form of a graph where the vertices are the elements of the system, and their complex interaction is put in the form of an edge (directed, weighted, etc).4 By considering the individual variability within patients, network theory makes it possible to optimize and individualize the diagnosis in the modelling of precision medicine.5,6

In the last decades, the evolution of disparate, complex biological
systems has been predicted, almost automatically, by using different tools and procedures from Artificial Intelligence (AI).

From an initial data set of patient’s characteristics ('learning set'), AI algorithms learn how the features relate to and predict the outcome. Machine Learning (ML), a sub-discipline in AI, is instead focussed on the ability to handle noisy or irrelevant data and on the capability to predict the outcome of a disease based on data derived from similar conditions. The first attempts to use biomedical data to extract prediction resulted in probabilistic models trained on a series of case studies, tried to match the individual patient’s condition with predefined classes of stratified increasing health risks.

Using Artificial Neural Network procedures, the correlation between early craniofacial features and the risk of craniofacial worsening during growth was established in 43 orthodontic patients. As the number of cases available for predicting health outcomes increased, new problems related to data incompleteness and homogeneity and ethical aspects of storage arose. Using Graph analysis, feature covariance across orthodontic patients allowed to disentangle Class III subjects from the topological patterns of other malocclusions. ML algorithms have been adopted to assist the orthodontist in treatment plan decision, including premolar extraction. Indeed, such essential decisions during orthodontic treatments can be subjected to different points of view, as they tend to be based on the practitioner’s experience and intuition. Resorting to ML procedures, one can decrease, at least in principle, personal biases in the treatment analysis and choices.

The precision of ML answers crucially depends on the quality of data input, both in terms of number and appropriateness. ML algorithms cannot work if provided with raw data, which must be transformed into domain-specific representative and salient information. Therefore, choosing the best characteristics (the most expressive of the problem) is crucial for predicting. It is also imperative to collect and give the machine data expressed as geometric, physiological, clinical and anamnestic parameters to make the most of what the current technology has to offer.

The integration of biomedicine and computer science is based on two concepts: Systems of the body have complex and dynamic biological properties that rely on the interaction of molecular agents sustaining the physiological functioning, as well as the pathogenesis of diseases at different scales; if complex interactions cannot be understood or processed by conventional methods, they can be investigated and explained using ML. The quantification of imaging biomarkers able to witness the status of a system of the body is the core business of medical research. Among the multiple choices available in the field, Magnetic Resonance Imaging (MRI) is the one that offers high accuracy and reliability at multiple spatial resolutions, with the invaluable plus to be non-invasive.

This work aimed to present an essential overview of ML techniques and possible applications in the orthodontic discipline, highlighting merits, potentials and potential improvements offered by the inclusion of networked multiscale musculoskeletal data congruently segregated for advanced AI procedures. The integration of biomedical databases and mathematical models in multiscale and multi-physics systems is a well-established research field that discusses dynamic and static mechanical dynamics as essential elements in constructing models of living organisms and diseases. Although it is well beyond the ultimate purpose of this paper, we feel confident that the inclusion of such modelling in the fusion of biomedicine and computer science presented here will be mandatory in the next future.

1.1 The orofacial system as a hierarchical multiscale complex system

Orthodontic researchers and practitioners are interested in extracting the most significant possible amount of information from all potential sources to provide the best diagnostic framing and treatment. It means mining significant patterns from past studies and tracing down the localization of the pathobiology underlying growing trajectories with and without unfavourable dento-skeletal growth. Such difference may point towards the prediction of modifiable skeletal disharmonies. However, it is problematic to understand a priori what would be the best data representation, and transform inputs into information that the machine could understand.

It is well recognized that bones and teeth structural and mechanical properties, and the elastic properties of muscles strongly influence the post-treatment stability.

Once conceptualized as a whole, bones, teeth and muscles represent a complex system whose physiology is related to their multiscale hierarchical structures and the precise organization of inorganic and organic phases at the nanoscale, microscale, up to the macroscopic scale. It is known how the effects of orthodontic treatments can lead to significant or poor results, depending on the period of growth during which it is implemented and on the quality of the bone and bone-tooth interface. These latter factors involve both the mineral part and the collagen of bone marrow. Therefore, it is crucial to intervene with orthodontic treatments in the most appropriate period of growth but only to know the characteristics of the cancellous bones. Currently, there are attempts to evaluate both the optimal period of treatment and the health of the cancellous bone related to vertebral maturation. The maturation of the different skeletal segments does not happen at the same time. For example, the upper and lower jaw complete their development at other times. Due to the quality of the bone, the teeth can move or rotate unexpectedly, even with a technically perfect orthodontic procedure. Parameters that quantify the upper and lower jaw spongy bone quality should constitute the dataset to be collected for an appropriate and inclusive treatment plan. It is worth quantifying MRI parameters such as T2* and Apparent Diffusion Coefficient (ADC) which are related to spongy bone quality. Efforts to translate new dental and bone microstructural investigations to clinical practice in orthodontics should be devoted to making reliable predictions about the evolution of dento-skeletal tissue. It is crucial to understand how the micro and nanoscopic rearrangement contribute to the macroscopic bone tissue modification.
Cells of the musculoskeletal system reside within complex and often interconnected 3D environments, which regulate musculoskeletal physiology and homeostasis (Figure 2). Maxillofacial development is also regulated by the surrounding soft tissues. Muscles affect skeletal components' size and shape with a possible different relationship between the muscular pressure and structural configuration in Class I, II, and III malocclusions. Several orthodontic studies suggest the usefulness of monitoring electromyographic (EMG) activity of the jaw muscles. Moreover, a close correlation between skeletal and muscle tissues states has been recently highlighted, and fat increase and decreased muscle function. Parameters such as MRI T2 and ADC have proven useful to evaluate the muscles' state because of the relationship between the microscopic structure and function.

As highlighted above, despite the multiscale hierarchical nature of the musculoskeletal tissue, only macroscopic cephalometric variables composed of linear and angular geometric measurements are considered in conventional clinical practice. However, due to the development of in vivo imaging technology, additional methods can be used to obtain a set of multimodal and multi-parametric measurements at different length scales, involving the microstructure, the topology of the craniofacial tissue, and also its physiology. In recent years, in addition to the conventional x-ray computer tomography (CT) and the cone-beam computer tomography (CBCT), MRI has shown great potential in dentistry. Furthermore, as most patients receive orthodontic treatment during childhood, further development of radiation-free techniques, such as MRI are highly desirable.
Together with the measurements provided by the imaging techniques, it is crucial to collect clinical and anamnestic data from patients. Some authors use genetic data to obtain a set of priors that could better supplement the inputs for correct patient classification. Modelling malocclusion progression means taking into account the complexity of the system. The complex interplay of causes...
behind atypical growth requires a different perspective about the disorders affecting the orofacial biological balance. By enlarging the basin of information about the system’s chemical, physical and mechanical properties, it is possible to enable its description and evolution using more in-depth ML and AI procedures. To achieve data more representative of the patient’s situation, and therefore, the most significant possible amount of information, different in vitro investigation techniques about orofacial tissue at various scale lengths could be of clinical and research interest (Figure 2). Once the most extensive possible spectrum of information has been obtained, deciding the best approach to combine the collected data are required. Here, we present different ML methods that stem from the complex network theory to offer the appropriate patients’ classification, without drawing on the choice of a specific and predictive biophysical model.

1.2 | Learning features from data

Biomedical data, such as growth data, are constantly evolving. The related information is disseminated in a network of interconnected pieces of local information (nodes), and memory is encoded in the topology and the strength of such multiple connections (edges), rather than stored in the single information as in statistical databases. The power of such networks resides in their capacity to learn. Both in living beings and computer algorithms, the strengthening of connections is the microscopic mechanism for elaborating the information. Artificial neural networks (ANN) use multiple layers of calculations to imitate the human brain's reasoning and draw conclusions from initial information. ANN can deal with complex intertwined problems (Figure 3). Different types of signs, symptoms, X-rays, risk factors, imaging results enter into ANN algorithms to find the most predictive combination of variables. ANN assign weights to some combination of nodes (the features chosen to ‘train’ the machine) to repeatedly optimize the model’s predictive performance. According to the number of layers included in the neural network, different machines can be realized, from simple machine-learning (ML) engines to ‘deep’ learning ones.

The canonical ML workflow involves four steps: (a) data gathering, cleaning and pre-processing, (b) feature extraction, (c) model training, (d) evaluation of results. Refining these steps can be complicated by intuition alone. The ability to separate patient’s outcomes requires more targeted diagnostics, as to subgrouping patients with significantly different growth trajectories and clinical courses despite the similar early diagnostic frameworks. Once ML approaches are matched to orthodontic data, challenges arise from data incompleteness, high-dimensionality, heterogeneity, dynamicity, sparseness and statistical noise that can be partially mitigated by Network representation. Due to such data’s complex and interconnected nature, any single model can deepen only a tiny part of the entire orthodontic domain. Whilst the conventional medical approach is based on the careful recruitment of clinical and laboratory data, testing of a diagnostic hypothesis, causes and effects of phenomena, significance, checking the initial hypothesis and so on, ML workflow focuses instead on the fast predictive performance of models and iterative improvement of the algorithms, coping with high-dimensional spaces, variability of features and formats.

1.2.1 | Data gathering and Feature transformation

Feature engineering (FE) is the process that transforms raw data into features to feed into the prediction models. Therefore, features sit between data and models in the ML workflow (Figure 4). Moreover, being interested in defining the quality of growth, in the beginning, it has to decide how to assign each patient to the proper category: as an example, favourable/unfavourable growth often represents the label of interest. Then, patients’ clinical and cephalometric data can be imported into the training data, together with the specific labels. It is the beginning of the supervised learning approach. The algorithm will find the function that links the patient’s input data with the outcome, minimizing the number of errors. However, raw data can rarely be entered without a preliminary selection to discard redundancies and simplify the system.

1.2.2 | Learning

Regularization is the crucial attribute for high-quality learning data from a large number of features. The process does not aim to predict labels within data of the learning dataset optimally but rather generalize the previous about new, previously unseen patients. Accordingly, data regularization is equivalent to imposing a penalty for the system’s complexity to improve performance. When the algorithm has few patient samples available from which it can learn the connection between the characteristics and the outcome, it ends up memorizing the training data set instead of learning general features of data (over-adaptation to data, a.k.a. ‘overfitting’). For this reason, a model will perform well on the training set and poor with new patient’s data. On the other hand, a small number of patient’s characteristics in the learning set provides an inadequate description of the problem at hand and may result in learning difficulties (‘underfitting’). Thus, there is a trade-off between the model performance and the size of the training set, that is, the patient coverage, the per cent of potentially eligible patients for that risk assessment can be completed.

Supervised ML learning is used when the output is known. The learning procedure primarily deals with regression and classification problems. It focuses on classification, which involves choosing between subgroups to describe a new patient best, and predictions, which consists of estimating an outcome of interest, such as the quality of future craniofacial growth.
In contrast, unsupervised learning is not intended to find outputs to predict, and it tries to find naturally occurring patterns within data. The patterns identified have to be evaluated for utility. Unsupervised models aim to discover groupings from data samples ‘x’ without knowing the label output ‘y’. The algorithm is provided with unclassified data records to recognize whether any existing latent patterns are present. Clustering, dimensionality reduction, like principal component analysis are leading examples of unsupervised learning approaches. The preliminary subdivision into subgroups of patients may lead to the subsequent estimation of specific risk factors. In particular, clustering refers to the extraction of group latent similarities within data that allow subjects to be grouped into subsections. The clustering assumption states that such subsections of subjects often exhibit the same outcome. In orthodontics, because of the complexity of medical data and heterogeneity of patients, identifying subsections of patients by intuition can be difficult.

Semi-supervised learning (SSL) is a mixture of supervised and unsupervised models. It analyses several unlabelled cases (patients) whilst augmenting its pattern recognition capacities with a small quantity of labelled data. This approach is promising in orthodontic longitudinal studies since it is laborious to find data that refer to many patients followed closely over time, at comparable regular intervals; even more, untreated patients followed longitudinally.

Finally, reinforcement learning (RL) is an approach where software agents take actions in an environment to maximize a cumulative reward. RL differs from SL in not needing labelled input/output pairs to be presented without requiring sub-optimal actions to be corrected. The goal is finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge).

### 1.3 Learning features from networks

Learning techniques rely on networked environments to conduct the learning process on data of interest. What if data are not generated from grid-like Euclidean structures (like images and videos) and represented as graphs with complex interdependencies between objects? For network-based methods, the learning procedure is performed by navigating in networks built from the input data according to some similarity criterion. As networks naturally contain topological information of data relationship, network-based methods take advantage of typical algorithms that use raw data. It must be stressed that network-based methods can be considered a general solution for learning tasks, even for data sets not represented by networks. One can apply network construction techniques to that data set to generate a network from the input data. Once the network is constructed, the learning process can be run. Patient stratification is a general target of network-based methods of machine learning. For example, cancer somatic mutation profiles are highly coupled with the biomolecular network, in fact, somatic mutations of a cancer driver gene may lead cancer genome evolution to mutations in other genes. Therefore, each patient with its somatic mutation profiles can be identified, and the similarity between patients can be used to mine for tumour stratification. Analogously, information about malocclusion within the orofacial system can be extrapolated from the correlation matrix of the orthodontic
features. The correlation matrix relates to a network whose vertices are associated with the features and the edges to their covariance across patients, allowing the visualization of malocclusion information (Figure 5A). The networks are obtained by fixing a threshold value \( T \): if the Pearson correlation between two vertices (features) is larger than \( T \), they are considered connected ("linked"). Several topological network metrics are typically considered to convey biological meaning to cephalometric correlations (Figure 5A), as ‘betweenness’ (a node centrality index) or modularity (the capacity of nodes to form communities characterized by similar properties). Once the graph is computed, the topological structure of the input is encoded in simpler structures (e.g., vectors of reals), and then the ML algorithms can be run.

The application of ML to graphs can be generally divided into two categories:

- Node-focused applications, where the final task is associated with specific properties of each node (examples include node clustering, link prediction, semi-supervised node classification).
- Graph-focused application, where the model to be realized is dependent on the whole graph structure (examples include estimation of properties of graphs and graph classification).

The machine-learning methods that operate on graph domain are known as Graph Neural Networks (GNNs). Due to its convincing performance, GNN has become widely applied to infer data described by graphs. In GNN, the concept of Node Embedding is implemented. It means to map nodes to a \( d \)-dimensional embedding space (low-dimensional space rather than the actual dimension of the graph) to embed close to each other similar nodes in the graph. The procedure aims to map nodes, so the similarity in the embedding space approximates similarity in the network. Training can be unsupervised,
semi-supervised or supervised. The supervised setting provides labelled data for training and is used for node classification. The semi-supervised setting gives a small number of labelled nodes and many unlabelled nodes for training. In the test phase, the transductive setting requires the model to predict the labels of the given unlabelled nodes, whilst the inductive setting provides new unlabelled nodes from the same distribution to infer. Most node and edge classification tasks are semi-supervised. The unsupervised setting only offers unlabelled data for the model to find patterns. Node clustering (features/patients grouping) is a typical unsupervised learning task.

1.4 | Perspectives

The parameters necessary to understand and predict the evolution of an orthodontic condition (such as a malocclusion) cannot be captured by a single measurement modality or using a few data type (as an example, geometrical data from cephalometrics, from CBCT or anamnestic data). The dental-musculoskeletal tissue complexity lies in the interconnection of morphological progression, functional and genetic data. Nowadays, new and more sophisticated AI algorithms can deepen these aspects. Unfortunately, they do not give a clue about the reasoning that led to deciding the saliency, the relevance or the causal importance of the variables. Interestingly, theoretical advancements with methods from statistical physics will also allow determining partial information as estimating longitudinal data from cross-sectional to create individual prognosis.

Humans are brilliant at clinical pattern recognition in dimensions equal to or less than three. However, most medical data's dimension is much higher than three, making cognitive analysis problematic or even impossible.

Multiple perspectives are required to study complex phenomena formed of parts in a non-random organization, in which our knowledge can only be partial and idealized. Forcing biological explanatory pluralism into the narrow computational framework appears to be a misleading strategy. The complexity of the musculoskeletal craniofacial system requires the use of novel sources of information. In vivo, MRI matches this framework optimally, providing the elucidation of geometrical and physiological tissue features, even at length scales smaller than voxel resolution of X-ray CT and CBCT in a radiation-free modality. The computational models applied to orthodontics must include the skeletal characteristics change over time (even at the end of growth) and the relationship between the mechanical properties of the complex system composed of bones and teeth and their multiscale hierarchical structures, whose precise organization orchestrates the whole physiology. The question is to what extent this new ‘deep’ reasoning leads to reliable and responsible decisions in the global growth of the skeletal bases and the maturation of the dentition. The course of events almost always deviates from what was predicted and planned. There may be areas of local densification of skeletal disharmonies that can act as growth attractors. These areas are difficult to detect clinically and neither be easily detected by computational systems.

The future of orthodontics inexorably will pass through a Copernican revolution that will lead algorithms to optimize therapies (‘personalized medicine’). Nonetheless, the process will not be complete until the complexity of the problem is fully addressed. The need for new imaging techniques (in vivo and in vitro) to characterize teeth, bones and muscles structures and physiology, the availability of genetic data, and the filtration of helpful information from them will be the route to a new vision of orthodontics.

ACKNOWLEDGMENT

Open Access funding provided by Universita degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. [Correction added on 27 May 2022, after first online publication: CRUI funding statement has been added.]

CONFLICT OF INTEREST

The authors have no conflict of interests to declare.

AUTHOR CONTRIBUTION

Conceptualization, TG, GDC; methodology, PA, SC.; writing—original draft preparation, TG, GDC; writing—review and editing, GC, AP; supervision, GC.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no data sets were generated or analysed during the current study.

REFERENCES

1. Andersen PW. More is different. Science. 1972;177(4047):393-396. https://doi.org/10.1126/science.177.4047.393
2. Polimeni A. Pediatric Dentistry (2nd Ed). Edra (Ed); 2020.
3. Yan G, Tsekenis G, Barzel B, Slotine JJ, Liu YY, Barabási AL. Spectrum of controlling and observing complex networks. Nat Phys. 2015;11:779-786. https://doi.org/10.1038/nphys3422
4. Scale-Free CG. Networks: Complex Webs in Nature and Technology. Oxford University Press; 2007. https://doi.org/10.1093/acprof:oso/9780199211517.001.0001
5. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. Nat Rev Genet. 2011;12(1):56-68. https://doi.org/10.1038/nrg2918
6. Holzinger A. Extravaganzai tutorial on hot ideas for interactive knowledge discovery and data mining in biomedical informatics. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Berlin: Springer; 2014. https://doi.org/10.1007/978-3-319-09891-3_46
7. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol. 2016;12(7):878. https://doi.org/10.15252/msb.20156651
8. Gili et al.  
9. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. Int J Med Inform. 2008;77(2):81-97. https://doi.org/10.1016/j.ijmedinf.2006.11.006
10. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6(1):26094. https://doi.org/10.1038/srep26094
11. DeoRC. Machine learning in medicine. Circulation. 2015;132(20):1920. https://doi.org/10.1161/CIRCULATIONAHA.115.001593
