AN EMPIRICAL APPROACH TO THE EXTRACTION VERSUS NON-EXTRACTION DECISION IN ORTHODONTICS

Mary Lanier Zaytoun

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the School of Dentistry (Orthodontics).

Chapel Hill
2019

Approved by:

Ching-Chang Ko

Sylvia Frazier-Bowers

Feng-Chang Lin

ABSTRACT

Mary Lanier Zaytoun: An Empirical Approach to the Extraction Vs. Non-extraction Decision in Orthodontics
(Under the direction of Ching-Chang Ko)

The extraction versus non-extraction decision is one of the most heavily debated topics in orthodontics. We hypothesize that orthodontic treatment planning can be enhanced by creating an empirical, evidence-based tool to aid in making this decision. To this end, we identified thresholds of overjet, overbite, and crowding that empirically determine whether or not to extract. The thresholds were combined into two prediction models: a decision tree and a logistic regression equation. These pilot models demonstrated clinical viability when tested against four borderline cases. To further improve these models, four additional models were built utilizing machine learning algorithms and an increased number of variables that influence the extraction decision, including demographics, additional clinical values, and cephalometric analysis. The best preforming model, with 81% prediction accuracy, was the convolutional neural network, which included 113 input variables. With continued development these models have potential for valuable clinical utility in the orthodontic profession.

ACKNOWLEDGEMENTS

Thank you to my committee members, Dr. Ko, Dr. Frazier-Bowers, and Dr. Feng-Chang Lin, for your expertise, guidance, and advice throughout my project. And a particularly special thank you to Dr. Ko for tirelessly guiding me in all of my research endeavors since dental school. Thank you to Dr. Feng-Chang Li, Yi Li, and Gloria Kim for the statistical analysis of the study data. Thank you to the Southern Association of Orthodontists for the generous research grant. Thank you to my family, for your selfless support and encouragement throughout all of my years of studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ROC	Receiver Operating Characteristic
CART	Classification and Regression Tree
CNN	Convolutional Neural Networking

AN EMPIRCAL APPROACH TO THE EXTRACTION VERSUS NON-EXTRACTION DECISION

Introduction

In orthodontics, the extraction of permanent teeth versus the expansion of dental arches to align teeth and correct malocclusion has been one of the most heavily debated topics in the field for decades. When deciding whether or not to extract permanent teeth, clinicians must weigh many factors including: spacing or crowding, overjet, overbite, occlusal stability, temporomandibular dysfunction, periodontal health, facial esthetics, smile arc, and systemic health among others. One of the factors that makes extractions one of the most difficult treatment approaches to prescribe is the fact that a significant proportion of orthodontic patients fit in the "borderline" category – meaning that extraction and non-extraction treatment plans can be considered equally. Historically, practitioners have learned to weigh all of the predisposing clinical factors and make a treatment decision by relying on their own clinical experiences and their training. The influence of each individual factor on extraction decisions has evolved over time as schools of thought have changed and new techniques and technologies have developed. Consequently, the rates of orthodontic extraction have also fluctuated over time.[1-9]

To evaluate this change in extraction rates over time, a forty-year study by Dr. Proffit analyzed the fluctuation of extraction frequencies at the University of North Carolina graduate orthodontic clinic.[4] Proffit found that the number of patients with extraction of all four first premolars increased from 10% in 1953 to 50% in 1963, remained between 35% to 45% until the early 1980s, then declined sharply back to the 10% level by 1993. He concluded that the initial increase in first premolar extractions occurred primarily in a search for greater long-term stability, while the recent decline seems due to a number of factors, namely: greater concern about the impact of extraction on facial esthetics, data to suggest that extraction does not guarantee stability, concern about temporomandibular dysfunction, and changes in technique.

A study by Janson et al. at a Brazilian University reported similar findings.[9] Their extraction rates peaked at 86% in 1973 and gradually fell to 46% in 2007. They attribute the decline in extraction rates to similar factors as reported in Proffit's study, as well as an increased utilization of growth modification approaches, the increase in frequency of maxillary expansion, and the increased usage of interproximal reduction.

In order to expand on this data and track extraction trends into the 21st century, Dr. Guez lead a follow up to Dr. Proffit's study, collecting the data of orthodontic patients treated at the University of North Carolina from the years 2000 to 2011.[10] Among this patient data set, they found that the factors leading to statistically significant odds of extraction ($p<.001$) were African-American race (as compared to the Caucasian reference group) as well as the clinical factors of initial overjet, overbite, maxillary crowding, and mandibular crowding. Their findings regarding rates of extraction during this time period were as follows: there was continued mild decreasing tendency towards extraction, both in overall extractions (leveling near 25% after 2006) and in four premolars extractions (just above 10%). This study also went beyond epidemiologic findings to determine the most statistically significant factors that influenced the clinician's decision to extract or expand. Guez identified the predisposing factors to be: ethnicity, Angle's Classification (skeletal and dental antero-posterior), overjet, overbite, and amount of crowding.

This paper seeks to expound on the studies by Proffit, Jansen, Guez, and many others, and advance the orthodontic extraction research to include an evidence-based prediction model for aiding in this highly debated clinical decision. Using the same database as Guez and collaborators, this study defines cutoffs for each clinical factor deemed significant in the extraction decision and proposes two evidence based algorithms that can be used by clinicians too make the extraction vs. non-extraction decision, especially in the cases on the “border-line” of extraction vs. non-extraction treatment plans. We hypothesize that a threshold value (cut-off) exists of overjet, overbite, and amount of crowding that when combined, can provide statistically-backed, evidenced based tools to assist in clinical decision-making. And, we also bring forward four borderline cases in which we demonstrate the effectiveness of these tools.

Materials and Methods

The aggregate data of 2,003 consecutively treated, orthodontic patients in the Graduate Clinic of the Orthodontic Department at the University of North Carolina from the 2000 to 2011 was analyzed. For each patient treated in this clinic, pretreatment variables are recorded and stored in a digital database. These records, which include demographic information, patient interview answers, and clinical exam findings, are standardized and stored in a secured digital database. Among the clinical exam findings are the factors that influence extraction rates. Records of consecutive patients were used in this study if their comprehensive orthodontic treatment began no earlier than January 1, 2000 and ended by December 31, 2011. Inclusion criteria were that the participant must have complete pre-treatment and post-treatment data present in the digital database. Third molar extraction was excluded from the outcome measurements for orthodontic tooth extraction. The outcome was the decision to extract teeth, or alternatively not to extract teeth, for orthodontic purposes. Approval for this study was given by the Institutional Review Board before data was gathered (IRB#132184).

Statistical Analysis

Receiver operating characteristic (ROC) analysis was used to evaluate the trade-off between sensitivity (true positive) and specificity (true negative) for the individual clinical variables of overjet, overbite, maxillary crowding, and mandibular crowding. ROC analysis operates under the principle that as sensitivity (the ability to detect true positive) is increased specificity (the ability to detect true negative) decreases. Therefore, in a series of clinically relevant values for each variable, sensitivity and specificity were calculated and ROC analysis was used to determine the optimal cutoff (threshold) of each variable by minimizing the difference between these two characteristics. Using an accrued score from the four dichotomized clinical variables (scoring one point for each observed value over the threshold), we developed a decision tree model that can be used to determine whether or not to extract teeth. In addition, we developed a logistic regression formula for calculation of the probability of tooth extraction using the measured clinical value (ie. amount of overjet in mm) of each variable.

Results

Using ROC analysis for the aggregate data of the 2,003 patients in this study, we estimated the threshold values for overbite, overjet, maxillary crowding, and mandibular crowding, as seen in Tables 1-4. By minimizing the difference between sensitivity and specificity, the optimal cut-off value for overbite was determined to be 3.5mm; the optimal cutoff value for overjet was determined to be 4.5mm; and the optimal cutoffs for crowding were determined to be 5.5mm in the mandibular arch and 6.5mm in the maxillary arch.

The composite score decision tree is seen in Figure 1. For each value that exceeds its respective threshold, a score of 1 is assigned. The sum of the scores for each factor results in a composite score. A composite score of less than or equal to 1 results in an extraction decision in 16% of the cases. A composite score of greater than or equal to 2 results in an extraction decision in 80% of the cases. Whereas a composite score equaling 2 results in an extraction decision in 45% of the cases. This result prompts a general decision rule that can be recommended in clinical practice – if there are 3 or 4 variables exceeding the threshold, extraction is recommended; if there are 0 or 1 variables exceeding the threshold, non-extraction is recommended.

The logistic regression formula for tooth extraction probability is shown in Figure 3. In cross tabulating extractions by the formula and actual extractions in our patient sample, the extraction cut-off was determined to be 36%, which results in 53% sensitivity and 86% specificity.

Cases

The aim of this study and these prediction models is to create a diagnostic, evidence-based aid in the decision that orthodontic clinicians must make on a day to day basis: whether to extract teeth or expand the arch to align teeth and obtain ideal occlusion and esthetics. In order to demonstrate this method's potential clinical usefulness and viability, four patients currently in treatment in the University of North Carolina Graduate Orthodontic Clinic have been selected as examples. These four patients represent borderline cases in which practitioners typically differ in their propensity for extractions.

Case 1 – A.R.

Patient A.R., a 15-year-old Hispanic male, presented to the orthodontic clinic with moderate to severe crowding and a chief complaint of "I don't like the appearance of my teeth. The bottom ones are crooked and the top canines stick out." He has skeletal and dental Class I relationships with 5mm overjet and 4mm overbite. He has a Brodie bite of his UL4. And his mandibular midline is mildly deviated from midsagittal plane.

A.R. is an example of a borderline extraction case. His maxillary canines are erupting ectopically and he has moderate to severe mandibular crowding.

Ultimately, due to the amount of crowding present in the maxillary and mandibular arches, the decision was made to extract maxillary and mandibular first premolars. Treatment included bonding full fixed appliances, self-ligating brackets with an .022 slot, and utilizing sliding mechanics.

Using our evidenced based models to confirm or deny or treatment decision - Patient A.R.'s composite score was calculated to be a 2. Given the composite score decision tree, there is a 45% chance for extraction in this case. And the logistic regression formula, with a p of 45%, indicates extraction. In this case, our clinical decision was confirmed by our predictive models.

Case 2 – S.B.

Patient S.B., a 13 year old African American female, presents to the orthodontic clinic with mild maxillary and mandibular crowding and a chief complaint of "I want to fix the crowding in the top teeth." She has Class I dental and skeletal relationships with bi-maxillary dentoalveoar protrusion. She presents with 3mm of overjet and 1mm overbite.

Given the vertical concerns with a decreased overbite, the mild to moderate crowding, and the existing dentoalveolar protrusion, S.B. is an example of a "borderline" case - arguments can be made to either extract teeth or to expand the arch.

Clinically, in S.B.'s case, the decision was made to proceed with this case non-extraction. The esthetics related to dentoalveolar protrusion were discussed with the patient and her parents and they expressed no concerns in this regard. Six months into treatment, after alignment, protrusion and esthetics were re-assessed with the patient and parents and all agreed to proceed without extracting. Interproximal reduction in the lower anterior segment has been utilized to aid in controlling vertical dimension and maintaining appropriate positive overbite.

Using our prediction models to confirm or deny our treatment decision - Patient S.B.'s composite score was calculated to be a 1. Given the composite score decision tree, there is a 16% chance for extraction in this case. And the logistic regression formula, with a p of 33%, also indicates non-extraction. In this case, our prediction models confirm our clinical decision.

Case 3 – L.O.

Patient L.O., a 12 year old Caucasian female, presents with moderate crowding and a chief complaint of "I don't like my crooked teeth in the front." She has a dental Class I relationship and mild Skeletal Class II skeletal, due to mandibular retrognathism, with 4mm overjet and 4mm overbite.

Given the amount of mandibular crowding that L.O. displays, the argument can be made to extract teeth in this case. However, clinically we decided to treat L.O.'s case non-extraction. Factors that weighed heavily into our decision to treat L.O. non-extraction include: retrusive and relatively retroclined presentation of her lower incisors, an increased gingival display above her central incisor on smile, and well-interdigitated buccal segments. Treatment is progressing well and there have been no significant negative side effects associated with this course of treatment.

Referring to our predictive models to confirm this treatment decision - Patient L.O.'s composite score was calculated to be a 0. Given the composite score decision tree, there is a 16% chance for extraction in this case. And the logistic regression formula, with a p of 11%, does not indicate extraction. This is another example of a case in which the predictive models confirmed our clinical decision.

Case 4 – S.S.

Patient S.S., a 16 year old African American female, presents with moderate to severe crowding and a chief complaint of "I don't like to smile because my teeth are so crooked." She has dental Class I relationship and a mild skeletal Class II relationship, maxillary prognathic, with 5mm overjet and 4mm overbite. She has a single tooth dental crossbite in the left anterior region (UL2 to LL3) and a mild mandibular midline deviation from midsagittal plane.

While arguments to extract or expand can be made in this case, our clinical decision was to extract four premolars. The most relevant clinical factor leading to this decision was the severe mandibular crowding. Other factors supporting this decision were the slightly protrusive presentation of the upper incisors due to mild maxillary prognathism, dental protrusion of the lower incisors, and the decreased incisal display on smile. Treatment is progressing well, buccal segments have maintained Cl I occlusion, and there have been no significant side effects with this course of treatment.

Utilizing the predictive models to confirm or deny or treatment decision - Patient S.S.'s composite score was calculated to be a 3. Given the composite score decision tree, there is a 80%

chance for extraction in this case. And the logistic regression formula, with a p of 83%, indicates extraction.

Discussion

Extraction verses non-extraction treatment decisions can be one of the hardest day to day decisions in a clinical orthodontic practice, especially when it comes to "borderline cases." As seen in the four borderline cases presented, both tools accurately predicted the treatment modality chosen. This indicates that this decision tree model and logistic regression formula can serve as a helpful evidence-based aid in predicting the best course of treatment when extractions vs. arch expansion is in question.

The strength of these models lie in the fact that they were created from a patient pool of more than 2000 patients treated by approximately 25 different attending orthodontists using a variety of treatment modalities and have varying treatment philosophies. Additionally, the demographics of our patient population is representative of the overall US population, according to US census date from 2010.[11] Both of these factors suggest that these equations could be applied accurately and reliably to the patient pool in average US orthodontic practices and to widespread clinical orthodontic practice at large.

One limitation of this study is that this model does not account for soft tissue characteristics. As Drs. Proffit and Sarver have noted, over the last few decades orthodontic diagnosis has shifted from a sole evaluation of hard tissue to a greater incorporation of soft tissue esthetics and facial appearance. Outcome assessment has evolved from a pure focus on ideal occlusion to a larger, more global focus on ensuring esthetic treatment outcomes. The "soft tissue paradigm" as it has been termed, has become a primary focus of modern orthodontic treatment planning and outcome assessment.[12,13] The most current body of literature suggest that overall, in systematic reviews with long term follow-up, there is no significant change in soft tissue esthetics when comparing individuals treated with extraction to those treated non-extraction.[14,15] But this does not mean that in individual cases soft tissue is unaffected by the extraction vs. non-extraction decision. For example, several studies indicate that the extraction of four premolars has a negative impact on esthetics by resulting in a more retruded upper and lower lip.[16,17] Conversely, there can also be negative esthetic side effects associated with the expansion/non-extraction approach, in which this treatment results in excessively protrusive lips and a "too full" facial appearance. Based on this research, it is clear that clinicians need to accurately assess soft tissue esthetics and take lip protrusion and retrusion into account on a case by case basis when deciding whether or not to extract or expand.

A second, and similar, limitation, has to do with demographic data. In these models demographic data, most notably sex, age, and race, are not taken into account. Various studies have shown that different ethnic backgrounds and genders have different clinical presentations and preferences for ideal lip protrusion and facial esthetics.[18] For example, Chong et al. found that there is a difference in preference for lip protrusion among Caucasians and Chinese populations, with the Chinese judges less likely to accept protrusive lips as an acceptable and pleasing esthetic result.[19]

To account for these variances, new artificial intelligent algorithms may provide better predictions. Recently, Jung and Kim used the neural network machine learning approach to classify extraction vs. non-extraction in 156 Korean patients with 12 cephalometric features and 6 additional clinical variables. Jung and Kim found that the success rate of the model for predicting the proper treatment diagnosis, extraction vs. non-extraction, was 93%.[20] For the future development of our model, not only will measures of dental protrusion and retrusion, soft tissue profile, and ethnic and demographic data will be taken into account, but using a machine learning approach will allow us to account for a much larger database of variables. The ultimate goal of this model being to provide accurate and reliable personalized orthodontic care.[20]

Conclusion

The debate in orthodontics regarding whether to extract teeth vs. expand teeth is one of the toughest decisions that clinicians face on a day to day basis. There is a body of research studying the trends of extraction and expansion throughout the history of the profession, but no significant attempts at creating an evidence-based way to make this decision. This study was intended to move the research in such a direction. By creating a pilot model, both in equation and decision tree form, we sought to introduce a clinically available tool to aid in the extraction and expansion decision. A future study adding soft tissue parameters and demographic information to the model will help to add clinical relevance to this model.

Table 1. Overjet values. Cutoff of 4.5mm yields a sensitivity of 48% and specificity of 67%

Overjet (mm)	Sensitivity	1 – Specificity	-log(Sensitivity* Specificity)
1.50	0.895	0.915	1.119
2.50	0.757	0.747	0.718
3.50	0.621	0.500	0.508
4.50	**0.460**	**0.326**	**0.508**
5.50	0.304	0.194	0.611

Table 2. Overbite values. Cutoff of 3.5mm yields a sensitivity of 61% and specificity of 50%

Overbite (mm)	Sensitivity	1 - Specificity	-log(Sensitivity* Specificity)
1.50	.227	.177	.704
2.50	.394	.289	.553
3.50	**.606**	**.504**	**.522**
4.50	.768	.703	.642
5.50	.878	.851	.884

Table 3. Mandibular crowding values. Cutoff of 5.5mm yields a sensitivity of 49% and specificity of 86%

Mandibular Crowding (mm)	**Sensitivity**	**1 - Specificity**	**-log(Sensitivity* Specificity)**
3.50	.734	.442	0.893
4.50	**.615**	**.271**	**0.802**
5.50	.486	.138	0.869
6.50	.347	.077	1.138
7.50	.216	.034	1.566

Table 4. Maxillary Crowding values. Cutoff of 6.5mm yields a sensitivity of 40% and specificity of 91%

Maxillary Crowding (mm)	**Sensitivity**	**1 - Specificity**	**-log(Sensitivity* Specificity)**
3.50	.705	.473	0.990
4.50	.603	.315	0.883
5.50	**.497**	**.166**	**0.881**
6.50	.397	.090	1.019
7.50	.277	.057	1.344

Table 5. Extraction by clinical formula x Actual extraction cross tabulation. The optimal cutoff probability for extraction decisions is 36% (53% sensitivity and 86% specificity)

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Actual Extractions</th></tr>
<tr><th>No</th><th>Yes</th></tr>
<tr><th rowspan="2">Extraction by logistic regression formula</th><th>No</th><td>1218
86%</td><td>248
47%</td></tr>
<tr><th>Yes</th><td>206
14%</td><td>278
53%</td></tr>
</table>

Figure 1. Composite score decision tree model. One point is scored for each outcome and the sum is the composite score for that patient.

Composite Score Calculation	
Overjet	> 4.5mm
Overbite	< 3.5mm
Maxillary crowding	> 6.5mm
Mandibular crowding	> 5.5mm

* for each clinical factor scoring >/< the indicated value, a score of 1 is assigned. The sum of the scores for each clinical factor = the composite score.

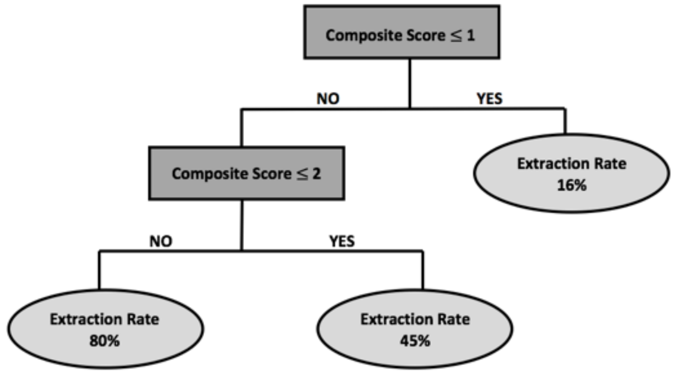

Figure 2. Logistic regression formula for extraction probability

$$\log\left(\frac{p}{1-p}\right) = -1.6 + 0.2(\textit{overjet in mm}) - 0.2\,(\textit{overbite in mm}) + 0.1\,(\textit{max crowding in mm}) + 0.1\,(\textit{mand crowding in mm})$$

Figure 3. Patient A.R. intraoral photos

Figure 4. Patient A.R., clinical data, composite score, and logistic regression outcomes

Overjet (mm)	Overbite (mm)	Maxillary Crowding (mm)	Mandibular Crowding (mm)	Composite Score	Extraction Probability (logistic regression formula)	Model Prediction	Clinical Decision
5	4	6	6	2	45%	Extract	Extract U/L4s

$$log\left(\frac{p}{1-p}\right) = -1.6 + 0.2\,(5mm) - 0.2\,(4mm) + 0.1\,(6mm) + 0.1\,(6mm)$$

$$log\left(\frac{p}{1-p}\right) = -0.20$$

$$p = 45\%$$

Composite Score Calculation		Score 1 or 0*
Overjet	> 4.5mm	1
Overbite	< 3.5mm	0
Maxillary Crowding	> 6.5mm	0
Mandibular Crowding	> 5.5mm	1
	SUM	2

*1 indicates threshold is met or exceeded for that factor.
0 indicates threshold is not met or exceeded for that factor.

Figure 5. Patient S.B. intraoral photos

Figure 6. Patient S.B.'s clinical data, composite score, and logistic regression outcomes.

Overjet (mm)	Overbite (mm)	Maxillary Crowding (mm)	Mandibular Crowding (mm)	Composite Score	Extraction Probability (logistic regression formula)	Model Prediction	Clinical Decision
3	1	1	4	1	33%	Non-extraction	Non-extraction

$$log\left(\frac{p}{1-p}\right) = -1.6 + 0.2\,(3mm) - 0.2\,(1mm) + 0.1\,(1mm) + 0.1\,(4mm)$$

$$log\left(\frac{p}{1-p}\right) = -0.70$$

$$p = 33\%$$

Composite Score Calculation		Score 1 or 0*
Overjet	> 4.5mm	0
Overbite	< 3.5mm	1
Maxillary Crowding	> 6.5mm	0
Mandibular Crowding	> 5.5mm	0
	SUM	1

*1 indicates threshold is met or exceeded for that factor.
0 indicates threshold is not met or exceeded for that factor.

Figure 7. Patient L.O. intraoral photos

Figure 8. Patient S.S. Clinical data, composite score, and logistic regression outcomes

Overjet (mm)	Overbite (mm)	Maxillary Crowding (mm)	Mandibular Crowding (mm)	Composite Score	Extraction Probability (logistic regression formula)	Model Prediction	Clinical Decision
5	4	4	5	1	38%	Non-extraction	Non-extraction

$$log\left(\frac{p}{1-p}\right) = -1.6 + 0.2(5mm) - 0.2\,(4mm) + 0.1\,(4mm) + 0.1\,(5mm)$$

$$log\left(\frac{p}{1-p}\right) = -0.50$$

$$p = 38\%$$

Composite Score Calculation		Score 1 or 0*
Overjet	> 4.5mm	1
Overbite	< 3.5mm	0
Maxillary Crowding	> 6.5mm	0
Mandibular Crowding	> 5.5mm	0
	SUM	1

*1 indicates threshold is met or exceeded for that factor.
0 indicates threshold is not met or exceeded for that factor.

Figure 9. Patient S.S. introral photos

Figure 10. Patient S.S.'s clinical data, composite score, and logistic regression outcomes

Overjet (mm)	Overbite (mm)	Maxillary Crowding (mm)	Mandibular Crowding (mm)	Composite Score	Extraction Probability	Model Prediction	Extraction Decision
5	4	7	9	3	55%	Extraction	Extract U4s/L5s

$$log\left(\frac{p}{1-p}\right) = -1.6 + 0.2(5mm) - 0.2\,(4mm) + 0.1\,(7mm) + 0.1\,(9mm)$$

$$log\left(\frac{p}{1-p}\right) = 0.20$$

$$p = 55\%$$

Composite Score Calculation		Score 1 or 0*
Overjet	> 4.5mm	1
Overbite	< 3.5mm	0
Maxillary Crowding	> 6.5mm	1
Mandibular Crowding	> 5.5mm	1
	SUM	3

*1 indicates threshold is met or exceeded for that factor.
0 indicates threshold is not met or exceeded for that factor.

REFERENCES

1. Hahn, G. W. (1944). Orthodontics: Its objectives, past and present. *American Journal of Orthodontics and Oral Surgery*, *30*(8), 401–404.

2. McLaughlin, R. P., & Bennett, J. C. (2015). Evolution of treatment mechanics and contemporary appliance design in orthodontics: A 40-year perspective. *American Journal of Orthodontics and Dentofacial Orthopedics : Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *147*(6), 654–662.

3. Konstantonis, D., Anthopoulou, C., & Makou, M. (2013). Extraction decision and identification of treatment predictors in Class I malocclusions. *Progress in Orthodontics*, *14,* 47.

4. Proffit, W. R. (1994). Forty-year review of extraction frequencies at a university orthodontic clinic. *The Angle Orthodontist*, *64*(6), 407–414. https://doi.org/10.1043/0003-3219(1994)064<0407:FROEFA>2.0.CO;2

5. Peck, S., & Peck, H. (1979). Frequency of tooth extraction in orthodontic treatment. *American Journal of Orthodontics*, *76*(5), 491–496.

6. Bramante, M. A. (1990). Controversies in orthodontics. *Dental Clinics of North America*, *34*(1), 91–102.

7. Baumrind, S., Korn, E. L., Boyd, R. L., & Maxwell, R. (1996). The decision to extract: part II. Analysis of clinicians' stated reasons for extraction. *American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *109*(4), 393–402.

8. Janson, G., Maria, F. R. T., & Bombonatti, R. (2014). Frequency evaluation of different extraction protocols in orthodontic treatment during 35 years. *Progress in Orthodontics*, *15*, 51.

9. Ackerman, J. L., Proffit, W. R., & Sarver, D. M. (1999). The emerging soft tissue paradigm in orthodontic diagnosis and treatment planning. *Clinical Orthodontics and Research*, *2*(2), 49–52.

10. Sarver, D. M. (9AD). Interactions of hard tissues, soft tissues, and growth over time, and their impact on orthodontic diagnosis and treatment planning. *American Journal of Orthodontics and Dentofacial Orthopedics*, *148*(3), 380–386. https://doi.org/10.1016/j.ajodo.2015.04.030

11. O'Connor, B. M. (1993). Contemporary trends in orthodontic practice: a national survey. *American Journal of Orthodontics and Dentofacial Orthopedics : Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *103*(2), 163–170. https://doi.org/10.1016/S0889-5406(05)81766-4

12. Jackson, T. H., Guez, C., Lin, F. C., Proffit, W. R., & Ko, C. C. (2017). Extraction frequencies at a university orthodontic clinic in the 21st century: Demographic and diagnostic

factors affecting the likelihood of extraction. *American Journal of Orthodontics and Dentofacial Orthopedics*, *151*(3), 456–462. https://doi.org/10.1016/j.ajodo.2016.08.021

13. QuickFacts. *U.S. Census Bureau QuickFacts: United States*, www.census.gov/quickfacts/fact/table/US/PST045216.

14. Rathod, A. B., Araujo, E., Vaden, J. L., Behrents, R. G., & Oliver, D. R. (2015). Extraction vs no treatment: Long-term facial profile changes. *American Journal of Orthodontics and Dentofacial Orthopedics*, *147*(5), 596–603. https://doi.org/https://doi.org/10.1016/j.ajodo.2015.01.018

15. Iared, W., Koga da Silva, E. M., Iared, W., & Rufino Macedo, C. (2017). Esthetic perception of changes in facial profile resulting from orthodontic treatment with extraction of premolars: A systematic review. *The Journal of the American Dental Association*, *148*(1), 9–16. https://doi.org/https://doi.org/10.1016/j.adaj.2016.09.004

16. Germeç, D., & Taner, T. U. (2008). Effects of extraction and nonextraction therapy with air-rotor stripping on facial esthetics in postadolescent borderline patients. *American Journal of Orthodontics and Dentofacial Orthopedics*, *133*(4), 539–549. https://doi.org/10.1016/j.ajodo.2006.04.052

17. Kouli, A., Papagiannis, A., Konstantoni, N., Halazonetis, D. J., & Konstantonis, D. (2018). A geometric morphometric evaluation of hard and soft tissue profile changes in borderline extraction versus non-extraction patients. *European Journal of Orthodontics*, 1–9. https://doi.org/10.1093/ejo/cjy056

18. Nomura, M., Motegi, E., Hatch, J. P., Gakunga, P. T., Ng'ang'a, P. M., Rugh, J. D., & Yamaguchi, H. (2018). Esthetic preferences of European American, Hispanic American, Japanese, and African judges for soft-tissue profiles. *American Journal of Orthodontics and Dentofacial Orthopedics*, *135*(4), S87–S95. https://doi.org/10.1016/j.ajodo.2008.02.019

19. Chong, H. T., Thea, K. W., Descallar, J., Chen, Y., Dalci, O., Wong, R., & Darendeliler, M. A. (2013). Comparison of White and Chinese perception of esthetic Chinese lip position. *The Angle Orthodontist*, *84*(2), 246–253. https://doi.org/10.2319/031213-212.1

20. Jung S-K and Kim T-W. (2016) New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop. 149:127-33.

MACHINE LEARNING ALGORITHMS: A NEW APPROACH TO THE EXTRACTION DECSION

Introduction

The extraction of permanent teeth versus the expansion of the dental arch has long been a source of debate in orthodontic treatment planning.[1-6] In many borderline cases, both extraction and non-extraction treatment options are justifiable. And clinicians tend to have a strong preference for their chosen treatment modality, extraction or non-extraction, based on their training, clinical experience, and personal treatment philosophies. But little research has been completed that quantifies this decision and that determines on a case-by-case basis, from an empirical standpoint, which treatment modality is the best fit for each individual patient. In the age of personalized medicine that relies on evidence-based science, we are seeking to advance this decision-making process by eliminating subjectivity and personal bias in favor of an empirical, statistically based algorithm for aiding in treatment decisions.

A previous study by Guez et al. identified overbite, overjet, maxillary and mandibular crowding as the statistically most important clinical variables in the extraction decision.[7] We continued Guez et al.'s research by identifying thresholds (cut-offs) for these predisposing factors. If the threshold for each variable is met or exceeded, extraction is indicated. By maximizing the difference between sensitivity and specificity, the cut-offs were determined to be 4.5mm for overjet, 3.5mm for overbite, 6.5mm for maxillary crowding, and 5.5mm for mandibular crowding. We further built a dichotomous tree and a logistic regression formula that combined these factors into chairside clinical aids for making these decisions. For the decision tree, each value that met or exceeded the cutoff for that factor was assigned a score of 1 and the composite score was compiled by calculating the sum of the factors. A low chance of extraction (16%) was indicated when the composite score was less than 1.5 while high extraction probability (80%) was predicted when the composite score was greater than 2.5. For the logistic regression formula, in cross tabulating extractions by the formula and actual extractions in our patient sample, the extraction cut-off was determined to be 36%.

While we believe these models have some clinical utility, we recognize that the extraction decision is very complex, and the four factors analyzed are not the only variables that influence this decision. Clinicians frequently cite other factors that weigh heavily in their decision making such as smile esthetics and gingival display, periodontal health, TMJ function, dental and lip protrusion, age, gender, ethnicity, among many other factors.[4-6] We hypothesize that by using machine learning algorithms and big data to incorporate hundreds of variables we can improve upon the dichotomous tree and logistic regression models previously mentioned and increase the clinical efficacy of this research.

Machine learning is a branch of artificial intelligence that employs a variety of statistical and optimization techniques that allow computers to "learn" from past examples and to detect hard-to-discern patterns and relationships from complex data sets. There are countless machine learning approaches and algorithms, each with their various strengths and downsides. After preliminary statistics, in which we explored five different algorithms for accuracy and utility in the extraction decision, we chose to focus this study on the three methods that showed the most accuracy in this initial dataset. Those methods are: Classification and Regression Trees (CART), Random Forest, and Convolutional Neural Networking (CNN).

Machine learning is a rapidly growing field, with applications in every facet of our lives. In other dental applications, various machine learning approaches have been utilized for purposes such as the diagnosis of bisphosphonate-related osteonecrosis of the jaw and for the detection of caries.[9,10] They have also been utilized in the medical world, some examples of such uses are the diagnosis of cancer based on imaging of tumor biopsies, the detection of potential adverse drug reactions, assessing risk prediction for cardiovascular disease, among many other applications.[11, 12, 13] Just as these algorithms have been utilized for diagnosis in other areas of dentistry and medicine, we intend to utilize machine learning and big data to diagnose the extraction vs. non-extraction decision in orthodontics. In fact, there is precedent for using machine learning to approach the extraction decision. A previous study by Jung and Kim constructed four neural networking models for this decision using the data of 156 patients, treated by one clinician, with input data consisting of 12 variables from cephalometric analysis.[14] We seek to improve on this data by using a larger and more varied patient pool (we utilize almost 700 more patients and our patient sample was treated by 25 different clinicians with varying treatment modalities and philosophies) and more diversity in the input data (we utilize an average of 100 additional input variables in each of our models).

Materials and Methods

Dataset

The aggregate data of 842 consecutively treated, orthodontic patients in the graduate orthodontic clinic at the University of North Carolina from 2010 to 2013 was analyzed. Our patient population characteristics as well as many of the characteristics that are relevant in the extraction decision are listed in Table 1. For each patient treated in this clinic, pretreatment variables - include demographic information, patient interview answers, and clinical exam findings - were recorded, standardized, and stored in a digital database. Additionally, cephalometric analysis was performed on initial cephalometric radiographs for each these patients in Dolphin Imaging platform. Among the data gathered (demographics, clinical findings, and ceph analysis) are the factors that influence extraction rates.

Records of consecutive patients were used in this study if their comprehensive orthodontic treatment began no earlier than January 1, 2010 and ended by December 31, 2013. Inclusion criteria were that the participant must have complete pre-treatment and post-treatment data present in the digital database. The outcome measured was extractions for orthodontic purposes, other than third molars.

Approval for this study was given by the Institutional Review Board before data was gathered (IRB#132184).

Machine Learning

The machine learning algorithms utilized in this study are Classification and Regression Trees (CART), Random Forest, and Convolutional Neural Networking (CNN).

CART is a non-parametric modeling technique for regression and classification problems using a decision tree. The decision tree makes sequential, hierarchical decisions using multiple predictors to arrive at a prediction on the outcome variable. It uses a set of consecutive binary rules to divide the sample into several subsamples. There are different algorithms that can be used to determine the best split variable at each node to divide the parent node into two child nodes. The selection of the split variable and split point can be chosen either simultaneously or one step after another. When there is no significant split variable nor split point for further split, the split is stopped. Consecutive splitting eventually grows the whole tree. Stopping and so-called pruning prevents the decision tree from overfitting.

Random forest is a machine learning algorithm composed of a combination of random sampling and multiple decision trees. As described above, in each decision tree, training consists of a set of consecutive binary rules to divide the sample into several subsamples. Each tree is trained independently with a randomly selected subset of the training sample with a subset of predictors and is allowed to grow to fullest diversity without pruning. The final classification, or output, is determined by taking the most common predictions from the terminal ensemble of each tree.[15, 16]

Convolutional Neural Networking is a deep learning model patterned after the operation of neurons in the human brain. It can be used for classifying data, clustering data by similarity, and performing object recognition. The CNN algorithm is composed of three types of layers – an input layer, an output layer, and multiple hidden layers in between consisting of convolutional layers and pooling layers. The neural network receives an input and transforms it through the series of hidden layers. Each hidden layer consists of a set of neurons that are fully connected all the neurons in the previous layer, but function independently relative to the neurons on their respective layer. The hidden convolutional and pooling layers extract data or features. And then the output layer, a fully connected layer, maps the data of the preceding hidden layers into an output or classification.[17]

Models

Utilizing these machine learning algorithms four models were built. Each model incorporates a different machine learning algorithm and/or a different set of variables

Model 1 is a CART using the four measurable clinical variables that were deemed significant to the extraction decision in our previous study. Those input variables are: overjet, overbite, maxillary crowding, and mandibular crowding.

Model 2 is a CART incorporating eleven variables. These input variables for this model include the four variables previously mentioned in Model 1 (overjet, overbite, maxillary crowding, and mandibular crowding) as well as additional clinical data (gingival attachment, Curve of Spee, skeletal anteroposterior relationship, and Angle classification) and demographic data (sex, race, and age).

Model 3 is a random forest algorithm with 113 input variables. The input variables for this model expands included the variables in Model 2 as well as the addition of 102 cephalometric variables from the tracing of the initial cephalometric radiograph.

Model 4 is a convolutional neural network algorithm with 114 input variables. This model uses the same variables as Model 3 but applies a different machine algorithm.

Accuracy Analysis

Before building the machine learning algorithms, 90% of the data was randomly selected into a training set while the remaining 10% of the data became the testing set. The training set was used to generate the prediction model while the testing set was used to estimate its accuracy. Accuracy was calculated to determine how well each model is able to predict the clinical outcome (extraction or non-extraction) that was chosen for each patient. In order to calculate accuracy and avoid overfitting of the model, 10-fold cross-validation is performed. For 10-fold cross validation, the original sample is randomly partitioned into 10 equal-sized subsamples. Of the 10 subsamples, we use 9 of them to train the model and apply to the 10th subsample, the testing sample. Then we repeat 10 times, with each of the 10 subsamples used exactly once as the validation (or testing) data. The average accuracy resulting from each of these testing sets is determined to be the model accuracy. By continually tuning hyper-parameters in the cross-validation, the model can be generalized to an unseen data set with better accuracy results.

ROC analysis was also used to evaluate the trade-off between sensitivity and specificity for each model. A high sensitivity (true positive) indicates the models ability to identify those patients who received extraction treatment. While a high specificity (true negative) indicates the ability oof the model to identify those patients who received non-extraction treatment. Ideally, the model the best model would have both high sensitivity and high specificity.

Results

In order to understand the true utility of these models in a clinical setting, the ability of each modela to correctly predict the treatment modality, referred to as model "accuracy", must be determined. Through 10-fold cross validation the accuracy of each model was calculated. Model 1, CART with 4 variables, had an accuracy of 77%; Model 2, CART with 11 variables, had an accuracy of 80%; Model 3, Random Forest with 113 variables, had an accuracy of 78%; and Model 4, CNN with 113 variables, had the highest accuracy of 81%. The associated rates of sensitivity and specificity are seen in Table 2.

In addition to accuracy, the most important variables in the algorithm were identified for Model 3. Because random forest algorithms score each variable on the level of importance or weight within the algorithm, identification of the most important variables for making the extraction decision can be made. Therefore, for Model 3 we were able to determine which variables the algorithm weighted most highly when making its extraction vs. non-extraction decision. These variables were, in ranked order: (1) maxillary crowding, (2) mandibular crowding, (3) Sn-GoGn value, (4) Sn-GoGn deviation, (5) maxillary unit length.

The first two variables, maxillary crowding and mandibular crowding, were standard clinical measurements, recorded in millimeters. The next two variables, SN-GoGN value and deviation, are measures gathered from each initial cephalometric tracing. SN-GoGN is an angular measurement of the skeletal vertical position. It is formed by intersecting the Gonion-Gnathion plane with the Sella-Nasion line. And it reflects the degree of inclination of the mandible relative to the anterior cranial base. A small angular measurement indicates a low plane angle, or "short" face. A larger angular measurement indicates a high mandibular plane angle, associated with a "long" face. "SN-GoGn deviation" specifically refers to the degree deviation of that particular measurement from the identified cephalometric norms, as defined by previous extensive research on the topic.

The variable importance is unable to be ranked in CNN, because CNN is a "black box" algorithm – only allowing for the identification of input and output variables, while each layer in between is hidden. And these hidden layers are transformational, with the variable importance changing after each layer.

Discussion

Based on these models, CNN (Model 4) appears to be the best model, showing the most promise for being a clinically utilized tool, for the extraction vs. non-extraction decision. This model had the highest accuracy, at 81%, with sensitivity and specificity values that remained relatively high as well, at 62% and 66% respectively. However, it is worth noting that 81% accuracy is not clinically viable and is still only marginally improved from Model 1, which incorporated only 4 variables. At this rate of accuracy many "borderline cases", those in which extraction and non-extraction approaches can be justifiable, are not accurately and predictably identified.

Also, it is noteworthy that Model 2 had an accuracy that was higher than Model 3, which is the reverse of what would be expected. Generally, with increasing data, random forest algorithms "learn" and become better able to detect patterns. However, the increased data, 113 variables in Model 3 vs. 11 variables in Model 2, did not improve Model 3's accuracy. This is possibly explained by the fact that Model 3 has a very high specificity (96%) and a very low sensitivity (23%), indicating the conservative nature of this model – it overwhelmingly classifies patients into the non-extraction treatment category.

The rank of variable importance in the random forest algorithm of Model 3 empirically corroborates common clinical thinking that crowding, both maxillary and mandibular, is the most influential factor when deciding whether or not to extract teeth. It is interesting to note that the third and fourth most important variables, SN-GoGn value and deviation, are measure of the skeletal vertical position. This would indicate that skeletal vertical, especially the extreme dolichofacial or brachyfacial presentations, play a critical role in the algorithms decision to predict extractions. This is a noteworthy finding because the previous research sited in this study did not include vertical considerations in their most prevalent factors that were considered in this decision. The results of this algorithm's variable importance would suggest that practitioners should weigh skeletal vertical position even higher than anterior-posterior considerations when making their extraction decision.

In terms of our data set, a strength of these models lies in the fact that they were created from a patient pool of more than 800 patients treated by approximately 25 different attending orthodontists using a variety of treatment modalities and have varying treatment philosophies.

Additionally, the demographics of our patient population is representative of the overall US population, according to US census date from 2010. Both of these factors suggest that these equations could be applied accurately and reliably to the patient pool in average US orthodontic practices and to widespread clinical orthodontic practice at large.

In terms of the strength of the input variables - these models, particularly Models 3 and 4, are comprehensive in their breadth of data and information. They incorporate nearly every clinical and demographic factor that a clinician must consider when making their extraction decision. Incorporating all standard measurement of cephalometric tracing, including cephalometric soft tissue measurements, is a novel advance for machine learning algorithms in this field.

Machine learning is a relatively new and still very much growing field. As machine algorithms continue to develop and become more complex, these models, particularly CNN, will continue to improve. As these models continue to advance, and to further strengthen and increase their clinical relevance, the next step of this research would be to divide the patient pool into subgroups based on categories such as race, angle classification, and "borderline cases." In doing this, we would be able to determine the models accuracy in each category – for example, these models may have a significantly increased accuracy in predicting extraction vs. non-extraction treatment in Angle Class II cases. If this example were found to be true, this finding could have major clinical implications in the way clinicians are able to utilize these models in Class II cases.

Conclusion

The extraction vs. non-extraction decision is one of the most difficult clinical decision that orthodontists face on a daily basis. Using machine learning algorithms, particularly CNN, shows promise in advancing this decision making process from a subjective decision to an evidence-based decision. This study advanced the research in this field by increasing the breadth of evidence available – using the largest patient pool as well as the most clinical variables, and testing four different models. Further development of the algorithm is needed before the model can reach a place of optimal clinical utility.

Table 4. Participant Characteristics. Model 1 includes 4 variables (initial maxillary crowding, initial mandibular crowding, initial overbite, initial overjet). Model 2 includes all variables reported in this table.

*level of statistical significance set to p = 0.05

Variables in Model 2	Variables in Model 1	Characteristic	Overall median (IQR) or % n = 842	Non-extraction median (IQR) % n = 634	Extraction median (IQR) % n = 208	p-value*
Model 2	Model 1	Initial maxillary crowding (mm)	2 (-2 to 5)	2 (-2 to 4)	4 (1 to 7)	<0.001
		Initial mandibular crowding (mm)	3 (1 to 5)	3 (0 to 4)	5 (3 to 7)	<0.001
		Initial overbite (mm)	3 (2 to 5)	4 (2 to 5)	3 (2 to 4)	<0.001
		Initial overjet (mm)	4 (2 to 5)	3 (2 to 5)	4 (2 to 5)	0.017
		Sex (%)				
		Female	59.1	57.9	63.0	0.224
		Male	40.9	42.1	37.0	
		Race (%)				
		White	61.5	67.2	44.2	<0.001
		African American	15.7	14.8	18.3	
		Other	22.8	18.0	37.5	
		Age at start of treatment (y)	14.3 (12.9 to 17.1)	14.2 (12.9 to 17.1)	14.6 (13.1 to 16.7)	0.490
		Skeletal AP relationship (%)				

		Class I	53.7	56.0	46.6	0.005
		Class II	33.6	30.6	42.8	
		Class III	12.7	13.4	10.6	
		Angle Classification (%)				
		Class I	39.8	42.6	31.2	0.010
		Class II	51.1	49.2	56.7	
		Class III	9.1	8.2	12.0	
		Initial Curve of Spee (mm)	2 (2 to 3)	2 (1.3 to 3)	2 (2 to 3)	0.887
		Reduced Attached Gingiva (%)	17.1	16.4	19.2	0.405

Table 2. Model Type with Accuracy, Sensitivity, and Specificity.

	Model Type	# of Variables included	Accuracy	Sensitivity	Specificity
Model 1	CART	4	77%	60%	83%
Model 2	CART	11	80%	40%	93%
Model 3	Random Forest	113	78%	23%	96%
Model 4	CNN	113	81%	62%	66%

REFERENCES

1. Konstantonis, D., Anthopoulou, C., & Makou, M. (2013). Extraction decision and identification of treatment predictors in Class I malocclusions. *Progress in Orthodontics*, *14*, 47.

2. Proffit, W. R. (1994). Forty-year review of extraction frequencies at a university orthodontic clinic. *The Angle Orthodontist*, *64*(6), 407–414. https://doi.org/10.1043/0003-3219(1994)064<0407:FROEFA>2.0.CO;2

3. Peck, S., & Peck, H. (1979). Frequency of tooth extraction in orthodontic treatment. *American Journal of Orthodontics*, *76*(5), 491–496.

4. Bramante, M. A. (1990). Controversies in orthodontics. *Dental Clinics of North America*, *34*(1), 91–102.

5. Baumrind, S., Korn, E. L., Boyd, R. L., & Maxwell, R. (1996). The decision to extract: part II. Analysis of clinicians' stated reasons for extraction. *American Journal of Orthodontics and Dentofacial Orthopedics: Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *109*(4), 393–402.

6. Janson, G., Maria, F. R. T., & Bombonatti, R. (2014). Frequency evaluation of different extraction protocols in orthodontic treatment during 35 years. *Progress in Orthodontics*, *15*, 51.

7. Jackson, T. H., Guez, C., Lin, F. C., Proffit, W. R., & Ko, C. C. (2017). Extraction frequencies at a university orthodontic clinic in the 21st century: Demographic and diagnostic factors affecting the likelihood of extraction. *American Journal of Orthodontics and Dentofacial Orthopedics*, *151*(3), 456–462. https://doi.org/10.1016/j.ajodo.2016.08.021

8. Burrow, S. J. (2008). To extract or not to extract: A diagnostic decision, not a marketing decision. *American Journal of Orthodontics and Dentofacial Orthopedics*, *133*(3), 341–342. https://doi.org/https://doi.org/10.1016/j.ajodo.2007.11.016

9. Kim, D. W., Kim, H., Nam, W., Kim, H. J., & Cha, I. H. (2018). Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report. *Bone*, *116* (January), 207–214. https://doi.org/10.1016/j.bone.2018.04.020

10. Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2018). Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry*, *77*(July), 106–111. https://doi.org/10.1016/j.jdent.2018.07.015

11. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

12. McMaster, Christopher. "A Machine-Learning Algorithm to Optimise Automated Adverse Drug Reaction Detection from Clinical Coding." *Drug Safety*, 02/2019, doi:10.1007/s40264-018-00794-y.

13. Goldstein, Benjamin A et al. "Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges" *European heart journal* vol. 38,23 (2016): 1805-1814.

14. Jung, S. K., & Kim, T. W. (2016). New approach for the diagnosis of extractions with neural network machine learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, *149*(1), 127–133. https://doi.org/10.1016/j.ajodo.2015.07.030

15. Breiman, L. E. O. (2001). Random Forest (LeoBreiman), 5–32. https://doi.org/10.1023/A:1010933404324

16. Kavzoglu, T. (2017). Object-Oriented Random Forest for High Resolution Land Cover Mapping Using Quickbird-2 Imagery. *Handbook of Neural Computation*, 607–619. https://doi.org/10.1016/B978-0-12-811318-9.00033-8

17. Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9