AI Built on the Consensus of 63 Orthodontists Predicts Extraction Decisions with 82% Accuracy

Lana Obeidat[a], Shivam Mehta[b], Aditya Tadinada[c], Gaurav Sinha[d], Madhur Upadhyay[a]

Division of Orthodontics, University of Connecticut School of Dental Medicine, Farmington, CT

Introduction

- Extraction vs. non-extraction is orthodontics' longest debate; extraction rates vary 10-90%
- Prior AI (14 studies): 54-95% accuracy, 1-30 evaluators, mostly retrospective and university-based
- **Question:** can AI trained on broad expert consensus plan like orthodontists?

Materials and Methods

6,341 patient data samples

63 board-certified orthodontists (5+ years)

- Each rated **15 clinical features** and chose primary, alternative, and extraction pattern plans
- 5+ years matches the proficient stage of the Dreyfus skill model, when intuitive decision-making develops
- Hierarchical CatBoost models; 80/20 split, 5-fold cross-validation

Input: 15 clinical features

↓

Stage 1: Primary plan
Extraction / Non-extraction / Surgical

↓

Stage 2A
Extraction per quadrant
UR · UL · LR · LL

Stage 2B
Alternative plan?
If yes, which one

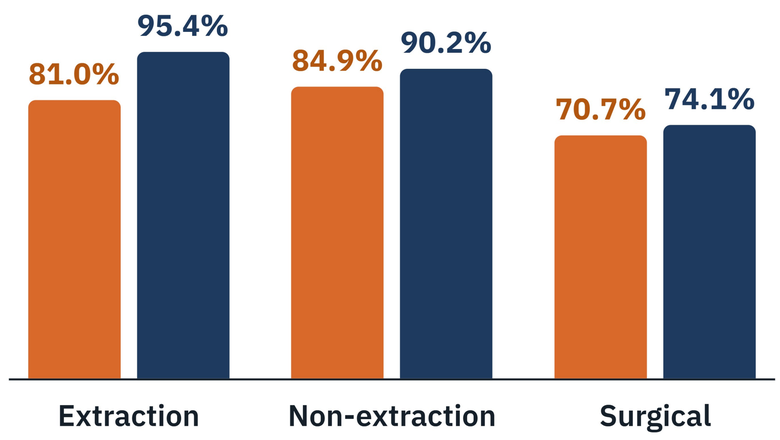
Results

82.2% primary plan matched experts (test set, n=1,269)

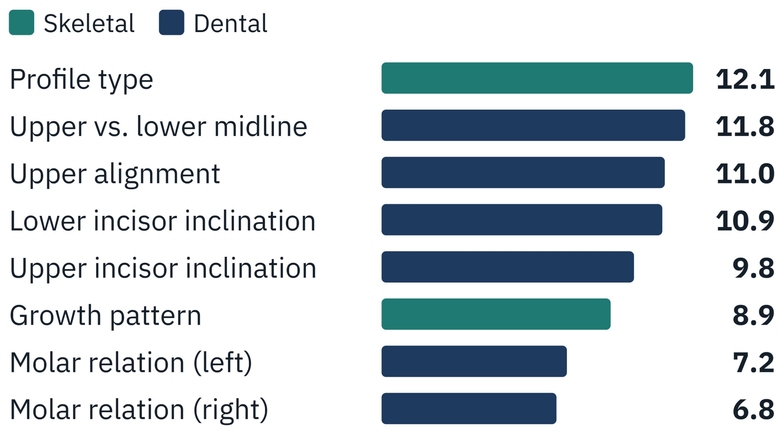
90.1% extraction cases with all 4 quadrants correct (613/680)

95.4% extraction cases matched a primary or alternative plan

Agreement with experts by treatment type



What drives the AI's primary treatment decision



Stage 1 feature importance, top 8 of 15 features. Profile type, midline, dental alignment, and incisor inclination are the most influential factors.

Compared with published AI systems

Prior studies: 1-30 evaluators

14 studies, accuracy **54-95%**

This study: 63 orthodontists accuracy **82.2%**

Largest dataset and expert panel reported to date

Conclusions

- Primary treatment plan predicted with 82.2% accuracy; strongest for extraction vs. non-extraction
- Extraction pattern predicted per quadrant with high accuracy (90.1% with all four correct)
- Surgical and alternative plans were harder to predict: fewer cases, more clinical subjectivity
- A hierarchical model that mirrors how orthodontists decide is practical for clinical use

Chairside role

- Verify a treatment decision
- Flag overlooked factors
- Add confidence in borderline cases

Limitations

What limited accuracy, and future directions

Limited surgical case representation (F1 51.2%)
→ Expand the surgical dataset and categorize surgical subtypes

Limited alternative treatment plans
→ Generate and rank multiple clinically acceptable options

No CBCT or tooth-specific health factors
→ Integrate CBCT, caries, restorability, and periodontal status

Lack of complete expert consensus
→ Evaluate the same cases across multiple experts; retrain from expert feedback

Literature cited

Baumrind S, et al. The decision to extract: Part 1. *Am J Orthod Dentofacial Orthop*. 1996;109:297-309.

Evrard A, et al. Which factors influence orthodontists in their decision to extract? *J Clin Exp Dent*. 2019;11:e432-8.

Tom K, et al. Crowding estimation and extraction recommendations. *Am J Orthod Dentofacial Orthop*. 2024;165:64-72.

Prokhorenkova L, et al. CatBoost: unbiased boosting with categorical features. *NeurIPS*. 2018;31.

Acknowledgments

The 63 participating orthodontists for their expertise, time, and dedication.

Supported by the NESO Resident Research Award (2025).

Further information

Lana Obeidat, DMD
lana.obeidat1@gmail.com

Conflicts of interest: None

Affiliations

[a] Division of Orthodontics, University of Connecticut School of Dental Medicine, Farmington, CT

[b] Department of Orthodontics, Texas A&M University School of Dentistry, Dallas, TX

[c] Department of Oral Health and Diagnostic Sciences, University of Connecticut School of Dental Medicine, Farmington, CT

[d] ProbAI Solutions