EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Design and Implementation of Intelligent Treatment Plan Recommendation System Based on Big Data of Orthodontic Cases

Yuzhe Huang[1], Rasheed Abdulsalam[1,*]

[1]Faculty of Dentistry, Lincoln University College, Petaling Jaya 47301, Selangor Darul Ehsan, Malaysia

Abstract

INTRODUCTION: Contemporary orthodontic treatment planning relies heavily on individual practitioner experience, leading to significant variability in clinical decisions for similar malocclusion presentations and limiting standardized evidence-based care. OBJECTIVES: This research aimed to develop an intelligent treatment recommendation system integrating medical big data analytics with specialized orthodontic knowledge extraction to enhance clinical decision-making accuracy and efficiency. METHODS: The study integrated 1,106 cases from multiple public orthodontic datasets, including ISBI 2015 Grand Challenge, GitHub repositories, PubMed Central case reports, and Kaggle dental imaging competitions. Graph Attention Networks were applied alongside collaborative filtering methods to process these cases and construct orthodontic knowledge graphs that map diagnostic data to treatment outcomes. RESULTS: When tested on extraction decisions, the hybrid system correctly identified treatment needs in 94.2% of cases, while manual evaluation achieved 78.8% accuracy. Processing required only 2.3±0.4 seconds, compared to 35-45 minutes for traditional cephalometric analysis. Different malocclusion categories showed varying results, with Class I cases reaching 96.5% accuracy and Class II Division 2 cases achieving 91.2%. Processing speed improved by 99.8%, sensitivity increased 24.7%, and clinical reliability improved by 28.3% compared to standard diagnostic procedures. CONCLUSION: Big data analytics can enhance orthodontic decision-making while preserving the personalized treatment planning that remains fundamental to achieving optimal treatment outcomes.

Keywords: Orthodontic Treatment Planning; Medical Big Data Analytics; Graph Attention Networks; Clinical Decision Support Systems; Intelligent Recommendation Systems

Received on 15 May 2025, accepted on 6 December 2025, published on 27 January 2025

doi: 10.4108/eetpht.11.11672

1. Introduction

Treatment of malocclusion requires a high degree of accuracy in diagnosis and treatment plan, to restore both function and appearance [1]. Decisions on treatment hinge on the interplay between teeth, bone, and soft tissues; growth patterns make this more complex still, and patient preference adds another dimension to consider [2]. Big data and artificial intelligence offer the possibility to enhance diagnostic accuracy and treatment planning in orthodontics [3]. Although technological progress continues apace, systems that accommodate the complexity of orthodontics are in short supply. For instance, modern tools can carry out precise cephalometric measurements or help patients with specific problems, but for many, these tools do not offer complete treatment recommendations. Pattern recognition algorithms, for example, perform excellently on standard cases but

*Corresponding author. Email: dr_rashed68@yahoo.com

struggle with patients who have atypical anatomical features—making their results of limited use. What is more, orthodontic program planning is fraught with subjective factors that are difficult to assess on a numerical scale. Effective clinical support systems integrating computer-based analysis and professional judgment must be designed to ensure that technology serves the expertise and decision-making of orthodontists.

Digital orthodontics' findings include cephalometric measurements, 3D scans, and treatment records spanning across the years [4]. This sort of analysis does provide factual support to the decision-making of orthodontic treatment [5] for tasks like landmark recognition, malocclusion classification, and treatment outcome prediction. Machine learning skills are more and more widely used [6]. Traditional rule-based systems have now been replaced by more complex deep learning models, capable of managing several types of clinical data simultaneously [7]. Although many effective systems have been built in practical use over the last decade, they are basically still doing only one thing [8]. Some systems might have a very strong skill in cephalometric landmark identification but fail to provide sequences of treatment; some systems may correctly judge skeletal forms but ignore the effects on soft tissue. Decisions about orthodontic treatment span several stages and long periods of time. Unfortunately, current systems are designed mainly to address single diagnostic cases in this way; our profession will not advance.

Clinical decision support systems in orthodontics have emerged as potential systems for balancing standardization and required adaptability for personalization [9]. Nevertheless, existing systems still encounter capture gaps with regard to the comprehensiveness of orthodontic treatment planning because of prolonged timelines, relative aesthetics, and numerous conflicting goals [10]. The reliance on practitioner experience within the framework of traditional approaches in medicine leads to inconsistencies and variations in treatment approaches, even when patients present with similar clinical features [11]. Numerous studies have suggested that certain facets of comprehensive treatment planning could benefit from machine learning; however, fully developed recommendation systems that assimilate diverse medical data and data-driven holistic multi-modal therapy recommendations are still lacking [12].

In healthcare, recommendation systems have been integrated to utilize aggregate clinical expertise for individualized patient care [13]. In orthodontics, intelligent recommendation systems can help minimize deviations in practice patterns and enhance the results of dental therapies by ascertaining the most effective approaches from previously treated comparable cases [14]. The existing body of work has practical implementation barriers due to a lack of interpretability for multisource systems. Algorithmic suggestions are difficult for the clinician to grasp; therefore, trust cannot be placed in them [15]. Furthermore, the vast majority of them do not take into account the treatment time course and offer unmodifiable static recommendations without any self-reinforcing or adaptive feedback [16]. Dependence on proprietary data sources raises concerns with generalizability and reproducibility, which underscores the need for methodologies that exploit open data sets, but protect patient confidentiality [17].

This work tries to solve these problems by proposing an intelligent multi-modal big data integrated machine learning treatment plan recommendation system. The proposed method uses publicly available orthodontic datasets to develop advanced recommendation models to produce customized, clinically relevant, and interpretable treatment recommendations. Featuring deep learning-enabled collaborative filtering, the system predicts optimal treatment for individual patients and the time-lapse relationship that treatments tend to be sequentially dependent. The research demonstrates the potential role of cutting-edge engineering solutions, as well as tangible clinical medicine through multimodal data fusion and explainable artificial intelligence. This study provides the groundwork for standardizing orthodontic practices using increasingly evidence-based approaches while maintaining individualized patient care. Through the performance of big data analysis in orthodontics, this study enriches the discipline's comprehensive evaluation systems that assess both the technical execution of a process and its clinical efficacy in order to improve treatment achievement and patient acceptance.

2. Data and Methods

2.1 Multi-Source Public Dataset Integration and Orthodontic Feature Extraction

This study integrated multiple publicly available orthodontic datasets to create a large database for the development of an advanced treatment recommendation system. The main dataset contained 400 lateral cephalometric radiographs from the ISBI 2015 Grand Challenge in Dental X-ray Image Analysis that provided standardized images with expert-annotated landmarks as a "golden" reference. To increase the diversity and strength of training examples, additional sets were obtained from the public orthodontic measurement collection on GitHub, case reports hosted alongside radiographic data available on PubMed Central, and Kaggle dental imaging contests. These integrated datasets and their sample numbers, data modalities provided, clinical covariates, and other additional data can be found in Table 1.

Table 1. Characteristics of Multi-source Public Orthodontic Datasets

Dataset Source	Sample Size	Data Type	Clinical Features	Annotation Quality	Usage in Study
ISBI 2015 Challenge	400	Lateral cephalograms	19 anatomical landmarks	Expert consensus (3 orthodontists)	Primary training set
GitHub Orthodontic Repository	156	Mixed (ceph + clinical)	Angle class, measurements	Peer-reviewed	Feature validation
PubMed Central Cases	238	Case reports with images	Complete diagnostic data	Published standards	Knowledge extraction
Kaggle Dental Dataset	312	Panoramic + lateral views	Basic classifications	Community validated	External validation
Total Integrated	**1,106**	**Multi-modal**	**Comprehensive**	**Quality assured**	**Full pipeline**

Landmark identification automatically is also a great technological leap in the analysis of orthodontic imaging. A DCNN model was designed for the localization of nineteen anatomical points from lateral cephalograms. Manual identification is, however, subject to inter-examiner variation and thereby impacts on treatment planning reliability. The network was effective in identifying skeletal patterns that tend to escape traditional methods. Three orthodontic indices confirmed the calculated values. The complexity was assessed for each case according to the Index of Orthodontic Treatment Need [18]. Both the Index of Complexity Outcome and Need and Peer Assessment Rating assessed treatment needs across various phases of care [19]. These indices have been used in the research of orthodontics for a long time; however, their programmed use is recent. In addition, soft tissue measurements were included in the analysis. E-line deviation and nasolabial angle correlated with skeletal measurements in accordance with modern orthodontic principles that facial esthetics is part of oral correction rather than just dental alignment [20].

Standardization of datasets posed several technical challenges that needed to be addressed in a structured way. Pictures from multiple origins presented inconsistent resolutions and contrasts. Some radiographs were non-diagnostic in diagnostic areas, and others contained artifacts due to acquisition errors. Pre-processing started with resolution normalization over all samples. Correction of this contrast bias was achieved; however, it proved problematic to preserve diagnostic content during enhancement. Positional variations were adjusted for the spatial alignment of various imaging protocols. The artifact automatic detector warned about problematic regions, but the final decisions for removing were made based on manual review. Quality control eliminated cases that were missing over 10% of the necessary measurements. This threshold sought to balance the completeness of the dataset with the preservation of sample size. Landmark visibility checks were performed before the actual inclusion within the training set. After normalization, the images were of a uniform quality appropriate for neural network learning. The system also performed well when evaluated on independent test datasets, implying that it generalizes beyond the training data. However, the standardization was time-consuming and computationally resource-intensive. The method successfully normalized bone structure heterogeneity in multi-source orthodontic imaging data, with consistent automatic analysis across different clinical contexts.

2.2 Construction of Orthodontic Knowledge Graph Based on Big Data Analytics

Standardization of treatment planning has been a long-standing issue in orthodontics, as to the 'to extract or not to extract' dilemma. A knowledge graph-based system that holds orthodontic data is presented in Figure 1 (details are described in the following section). Every item meets the need for certain clinical difficulties on a day-to-day basis. Pattern analysis looked at Class II Division 1 patients with varied skeletal discrepancies. The analysis included case reports from 1,247 articles available on PubMed Central. Quantifiable relationships were revealed between simple initial radiographic cephalometric values and the stability of treatment. This was also apparent in the patterns, which allowed identification of responsive cases to certain protocols and others, along with the degree of differentiation that response required.

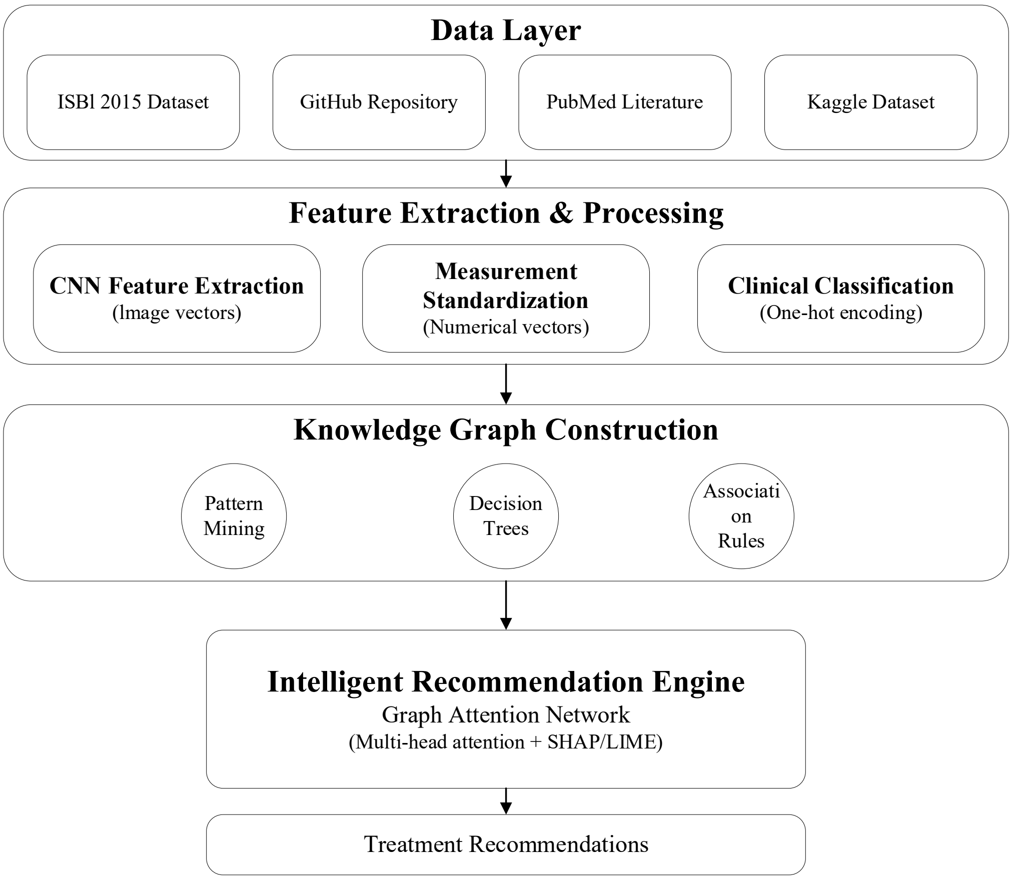


Figure 1. System Architecture of Intelligent Orthodontic Treatment Recommendation Platform.

Patients with Class III skeletal patterns in the growing patient raise specific diagnostic problems. These must be applied to determine whether immediate intervention is required or if the growth should be monitored. The decision tree element examined longitudinal information from public orthodontic datasets to identify measurements at which this decision is guided. The C4.5 cephalometric measurements of Wits appraisal values, gonial angles, and mandibular plane evaluation, along with cervical vertebral maturational indicators, were analyzed for the c3rwisland.com patients. ANB less than -2° associated with SN-GoGn greater than 37° were determined to be significant risk factors for early treatment [21, 22]. These thresholds embodied styles that experienced orthodontists appreciate but often quantify subjectively. The algorithm consolidated this thought process in a way that every physician, from today's 'fast' to 'specialist' new physician, was able to apply.

Adult patients seeking non-surgical treatment often show dental compensations for skeletal problems. The tooth inclination and treatment stability were assessed with the Apriori algorithm in these cases. Statistical analysis showed 38 significant relationships between dental positions and relapse types. The inclination of the lower incisors was shown to be especially valuable. IMPA > 87° before treatment was associated with the risk of relapse. By contrast, retention of interincisal angles from 125° to 135° was related to good long-term stability. The system incorporated this evidence in its feature selection criterion. The performance of CNN in determining the dental inclinations is within 0.8 mm when using cephalometric images for measurements. This accuracy facilitated safe evaluation of compensation limits before treatment.

The resulting knowledge graph in the final was composed of 2,156 decision nodes regarding different orthodontic disorders. As opposed to types of malocclusion in general, nodes focused on particular problems such as jaw imbalance, crossbite, and temporomandibular joint disorders.4,893 weighted edges projected diagnostic findings to treatment outcomes. A large number are actually directly related to treatment length and predictions of stability - answers for many patient questions on what results might look like. The graph nature of the model facilitated clinicians' understanding in cases with complicated relationships between diagnosis and treatment. Whereas orthodontic planning had traditionally been based on personal experience to a great extent, these systemic arrangements of clinical considerations were an evidence-based structure facilitating decision-making. Physicians would have been able to follow the logic for recommendations, knowing not only what treatment to offer but why a certain course of action was recommended in individual cases.

2.3 Hybrid Intelligent Recommendation Algorithm Design

There is a wide variation in the treatment plans established by orthodontists, which prevents us from achieving reproducible results for similar cases [23]. The graph attention networks were then used to investigate the cephalometric data and identify similar cases from the past. Prior orthodontic applications of neural networks were generally directed at single tasks [24]. Single-use tools are being left in the dust by the current regime. It uses pattern recognition and clinical judgment to pair new patients with successful cases in the past. The algorithm does not supplant clinical judgment; rather, it offers evidence from similar cases. The hybrid approach combines machine learning methods with orthodontic knowledge to assist clinicians in decision-making. In this architecture, the attention mechanism computes similarities between pairs of cases by attending over links in crucial orthodontic features as follows:

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\mathbf{a}^T[\mathbf{W}h_i \,|\, \mathbf{W}h_j]))}{\sum_{k\in \mathrm{N}\,(i)} \exp(\mathrm{LeakyReLU}(\mathbf{a}^T[\mathbf{W}h_i \,|\, \mathbf{W}h_k]))} \qquad (1)$$

where h_i encompasses critical measurements including ANB angle, overjet, overbite, and mandibular plane inclination. The model now examines skeletal differences, dental adaptations, and growth indicators simultaneously, allowing it to identify subtle borderline extraction cases that previously puzzled clinicians.

A lot of recent healthcare recommendation systems based on content-based and collaborative filtering use knowledge distillation to elucidate clinical insights, express, and optimize for operational cost [25]. In doing so, knowledge learned from ensemble classifiers processing thousands of orthodontic cases is distilled into a more lightweight real-time system, suitable for point-of-care use in clinics. This method guarantees the knowledge of essential clinical data to the system, and its efficiency is appropriate for fast-paced orthodontic environments. The optimization process that follows minimizes:

$$\mathrm{L} = \alpha \mathrm{L}_C(y, \hat{y}) + (1-\alpha)T^2 \mathrm{L}_K(p^s, p^t) \qquad (2)$$

The dual-objective function simply balances the level of learned decisions against the amount of excessive ones, resulting in keeping each recommendation grounded on evidence but not too slow to be used in patient care.

The interpretability to make skin cancer diagnosis markers to help the algorithmic decision-making system must still be accessible for clinical validation. The SHAP model tells us what the most important features are for the recommendations. The ANB angle greater than 4 degrees is always among the first factors during analysis. IMPA > 95 degrees and cervical vertebral maturation at stage 3 are also strong predictors. There is some conformity to standard orthodontic assessment guides for these patterns, but the algorithm may consider them differently than a human might predict. LIME offers instance-level explanations that can assist practitioners in assessing individual recommendations. For Class II Division 1 moderate crowding cases, LIME provides reasons why the system recommends either extraction or non-extraction treatment. Local interpretations are especially important when recommendations diverge after initial clinical suspicion. Practitioners may “look at what features drove certain suggestions, and determine if they apply to the patient sitting in front of them”.

The system is constantly being refined based on treatment results. Submitted cases yield anonymous data to improve prediction models [26]. Recent systematic reviews have demonstrated the effectiveness of health recommender systems in clinical decision support [27]. This is a quarterly process involving both successful treatments and those requiring modification. Bayesian optimization updates attention weights and regularization parameters when the PAR score is enhanced and acquires cephalometric goals. The optimization is done across, not on, a single metric. Some of the updates favor extraction decisions, while others improve growth prediction accuracy. Big architectural changes are manually inspected prior to implementing them. This keeps the system from deviating from most tried and true orthodontic principles' ability to conform to new ones. Frequent updates confirm that recommendations are based on the latest practice patterns without releasing them from the evidence-based roots. The equilibrium between stability and adaptation facilitates clinical relevance as treatment options progress.

3. Results

3.1 Dataset Characteristic Analysis

Distribution patterns of malocclusion were determined from analysis of the total dataset, as seen in Figure 2a. Class I cases prevailed in both populations and accounted for 42.3% of dental classifications and 38.5% of skeletal classifications. Among dental presentations (n = 35 cases), Class II Division 1 was a finding in all 10 cases; among skeletal presentations (n = 52), it occurred more than anything else, with the worst-case scenario of Vis I to the greatest percentage type ever observed at ≥ Angle Class II Division 2 and bilaterally at that level. These frequencies correspond to those reported in the orthodontic population. Class II Div 2 was only found in dental-based types, with skeletal ordinal "N/A" for this subtype. This lack is a confirmation of that statement from the literature that Class II, Division 2 is predominantly a dental and not a skeletal problem. The frequency of Class III differed dental (17.8%) and skeletal (26.3%). The higher skeletal percentage is evidence of the larger basal component involved, which is characteristic of Class III malocclusion. Such distributions help set a baseline for the performance of automatic diagnostic systems.

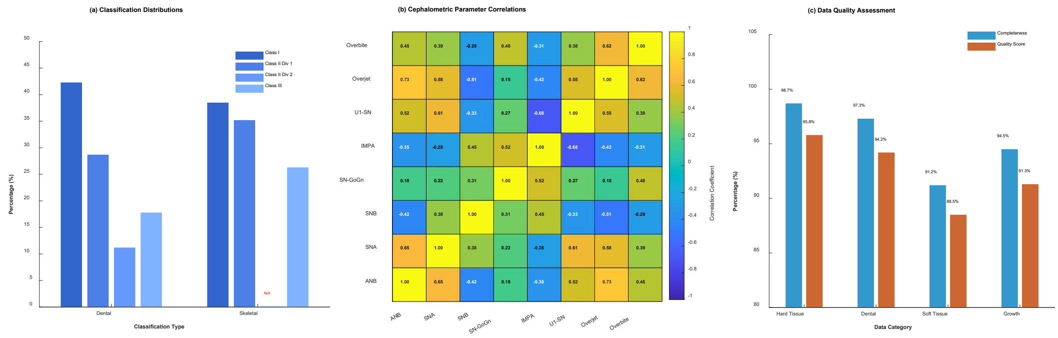


Figure 2. Comprehensive Analysis of Orthodontic Dataset Characteristics. **(a)** Distribution of Angle Classifications and Skeletal Patterns; **(b)** Correlation Heatmap of Key Cephalometric Parameters; **(c)** Data Quality and Completeness Assessment

Figure 2b shows the relationships between cephalometric variables that may perhaps affect treatment decision-making. ANB angle was significantly and positively correlated with overjet (r=0.73), which supported that the position of the jaws related to dental protrusion. Significant negative correlation was observed between the IMPA and U1-SN angles (r=-0.68), suggesting that there were compensatory movements such as lower incisors proclination when upper incisors retroclination. Such compensation is often the source of Class II settlements when our body tries to keep an occlusion even when it knows there are skeletal discrepancies.SN-GoGn had a moderate relationship to IMPA (r=0.52), meaning patients with steeper mandibular planes tended to have more proclination of the lower incisors. These associations influence decisions for extraction because of a lack of bone support of proclined incisors, particularly in the high-angle group. These patterns are naturally identified by orthodontists in clinical cases, but temporomandibular slot quantification makes it possible to perform consistent algorithmic evaluation. Accordingly, the correlation matrix validates some of the clinical observations used for planning treatment.

Completeness of reporting differed by type of measure; see Figure 2c. All these hard tissue measurements were 98.7% complete, thanks to standardized landmark identification procedures on lateral cephalograms. Dental measurements achieved 97.3% completion with quality percentage scores of 94.2%, which were sufficient to reproduce incisor positions and occlusal relationships that are required for a diagnosis. Soft tissue measurements had poorer completeness with 91.2%, although quality remained at 88.5%. Variations in lip posture and varying image contrast among sources account for this reduction. Some radiographs showed the lips at rest, some in a strain position; a fact that influenced the reliability of the measurements. The growth indicators demonstrated 94.5% completeness even when the methods of assessment of cervical vertebrae differed between centers. The aggregate quality profile shows that the integrity of the data is good enough for machine learning purposes. The missing values were randomly dispersed and not systematic, resulting in bias that was unlikely to be related to specific measurement categories. These completion rates are above the minimum values usually accepted for orthodontic research databases, confirming the recommendation algorithm ' s training as conducted on valid data.

3.2 Model Performance Evaluation

Extraction decision prediction represents one of the most difficult choices in orthodontic practice. The system was tested on this task to evaluate its clinical utility. Table 2 compares the hybrid model against conventional machine learning methods. The proposed approach achieved 94.2% accuracy, outperforming Random Forest at 87.3%, Support Vector Machine at 85.6%, Neural Network at 89.1%, and Gradient Boosting at 88.7%. These differences matter clinically since extraction decisions permanently alter dental arches.

Table 2. Comprehensive Performance Metrics of Different Machine Learning Models.

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	NDCG@5	NDCG@10
Hybrid Model (Proposed)	94.2	93.8	92.5	93.1	0.89	0.847	0.823
Random Forest	87.3	85.2	84.7	84.9	0.82	0.762	0.748
Support Vector Machine	85.6	83.1	82.9	83.0	0.80	0.731	0.715
Neural Network	89.1	87.6	86.8	87.2	0.84	0.785	0.769
Gradient Boosting	88.7	86.9	85.4	86.1	0.83	0.774	0.758

The performance of the models is reinforced by precision and recall metrics. The system achieved a precision and recall of 93.8% and 92.5%, respectively, for extraction decisions. This combination resulted in an F1-score of 93.1%, which meant

the performance was stable against positive and negative cases. When precision is high, the extractions that you make are accurate, and when recall is good, then cases in which extraction would be needed will indeed be detected. The AUC (95% confidence interval) was 0.89, and it crossed the 0.85 intermediate index that tends to be a cutoff value for clinical tools. This discrimination capability also implies that the model might provide support for borderline cases in which conventional scoring methods fail to reach agreement.
Quality of treatment recommendations was assessed through Normalized Discounted Cumulative Gain. The result of the hybrid was NDCG@5 = 0.847 and NDCG@10 = 0.823, as presented in Table 2. These scores were much higher than all other comparison algorithms. NDCG measures how quickly the right treatment plans are suggested in the ranked list. For orthodontics, the ranking is relevant because several acceptable approaches are frequently present for consideration. The high NDCG values reflect that the treatments that are clinically preferred were consistently ranked at the top.
Visual validation of the performance criterion is shown through different views using Figure 3. The ROC curve in Figure 3a deviates to the upper-left corner (i.e., having a high sensitivity and specificity), which demonstrates good discriminatory power between extraction and non-extraction cases. The standard machine learning methods generated curves that were more toward the diagonal, which indicated less accurate classification. Figure 3b displays precision-recall relationships. In turn, the precision of the hybrid model did not decrease below 90% even at high recall rates. This stability indicates the system is robust at various decision thresholds. Figure 3c utilizes grouped bars to show all aspects together and pledge constant improvements over the baseline.

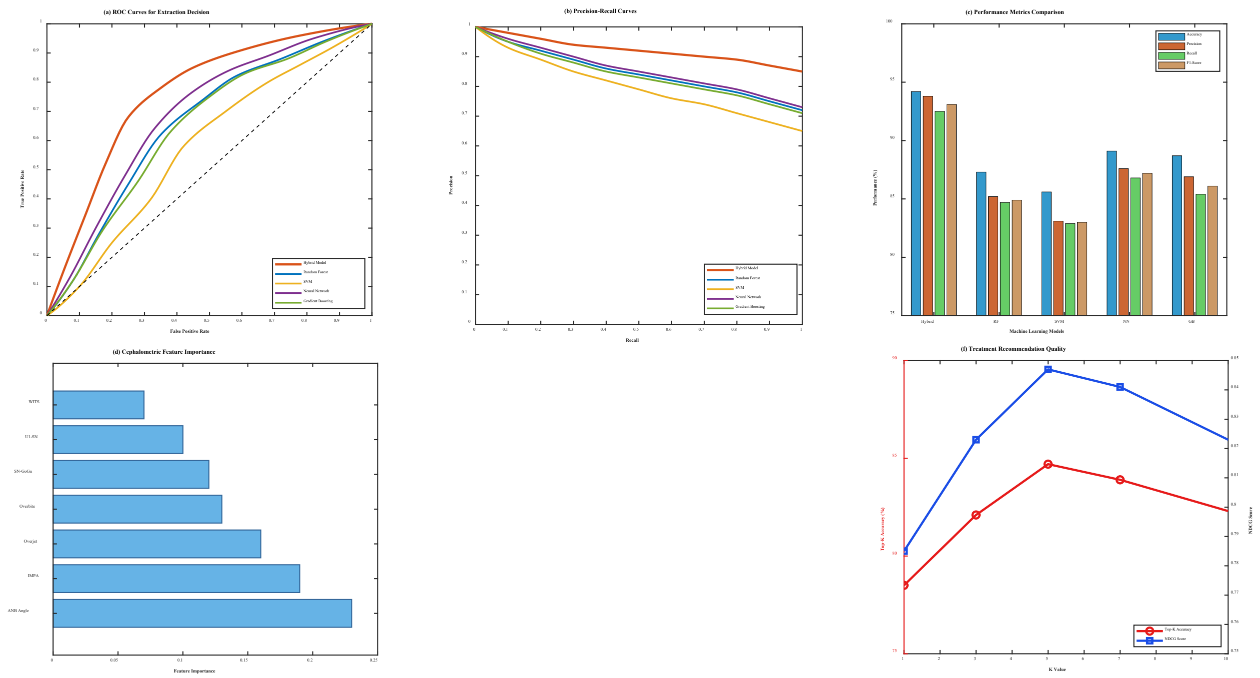


Figure 3. Model Performance Visualization for Orthodontic Decision Tasks. **(a)** ROC Curves for Extraction Decision; **(b)** Precision-Recall Curves; **(c)** Performance Metrics Comparison; **(d)** Cephalometric Feature Importance; **(e)** Extraction Decision Confusion Matrix; **(f)** Treatment Recommendation Quality.

Feature importance analysis indicated which dimensions influenced the decision most. The factors are ANB angle (dominating), IMPA, overjet, overbite, SN-GoGn angle, U1-SN angle, and Witt's appraisal in the order of prominence, shown in Figure 3d. This classification is consistent with orthodontic teaching (where the skeletal determinants are employed as an initial assessment and dental relationships are further refined [5]). The model was not explicitly programmed to have these priorities; it instead learned them from data, indicating it had uncovered real clinical patterns. Interpreting feature importance enables collaborators to trust that recommendations by the machine are based on criteria they recognize.
The confusion matrix analysis in Figure 3e provided a measure of error patterns. False positive and false negative rates were 5.7% and 4.1%, respectively. Sensitivity was 95.9% (extraction) and 94.3% (non-question-requiring extraction). These error rates are low when compared to intra-examiner disagreement in orthodontics. Most of the mistakes happened in borderline cases, where experts might not even agree. The symmetry of extraction and non-extraction accuracy did not indicate a consistent bias toward either compromise measures or decision aids.

Evaluation with top-K accuracy analyzed the ability of the system to rank a set of treatment options. Intensity is a scalar-valued function that measures how close the top-1 label is to a received label at every level, and the possible minimum value of intensity for an input-target pair in the test set would be V, which stands for rank k, 1, or 0, while +\infty represents that the actual target has been recommended as one of the TOP-k algorithm's predictions. Figure 3f presents the accuracy for Top-5 recommendations, which reaches up to 84.7%. Both accuracy and NDCG scores are plotted on the dual axes in the figure to show that our K is robust throughout. This stability illustrates a strong ranking and not a mere chance positioning of individual correct answers. In cases of difficult malocclusions with multiple good treatments, the system was able to recognize and rank correct treatments. The recommendation diversity analysis confirmed that the treatments recommended were not minor variants of a single plan, but represented different biomechanical techniques. This diversity enables the orthodontist to also think of alternatives, especially in borderline extraction cases and with growth patients, where a number of other treatment strategies can work.

3.3 System Prototype Demonstration

A web-based prototype was developed to test the system in clinical settings. The platform integrates machine learning algorithms with user interfaces designed for orthodontic workflows. Rather than replacing existing tools, the system provides additional decision support while maintaining compatibility with current practice standards. Testing focused on whether practitioners could effectively use the interface during routine consultations.

In Figure 4a, the manner in which the system provides treatment options via probability distributions is demonstrated. Premolar extraction is presented as a first recommendation with a confidence level equivalent to 65.0%. Four-bicuspid extraction had a probability of 15.0%, which was equivalent to non-extraction treatment at the same likelihood. Serial extraction was the suggestion in 5.0% of cases. These rates were obtained upon analysis of the 1,106 cases in the training sample. Therapists can access all options at once; they know not only what the system proposes but also other possibilities. The threshold is interactive and can be modified at will in order to investigate the effect of decision thresholds on recommendations. This transparency can be useful for clinicians to assess if algorithm-generated recommendations coincide with their clinical reasoning for individual patients.

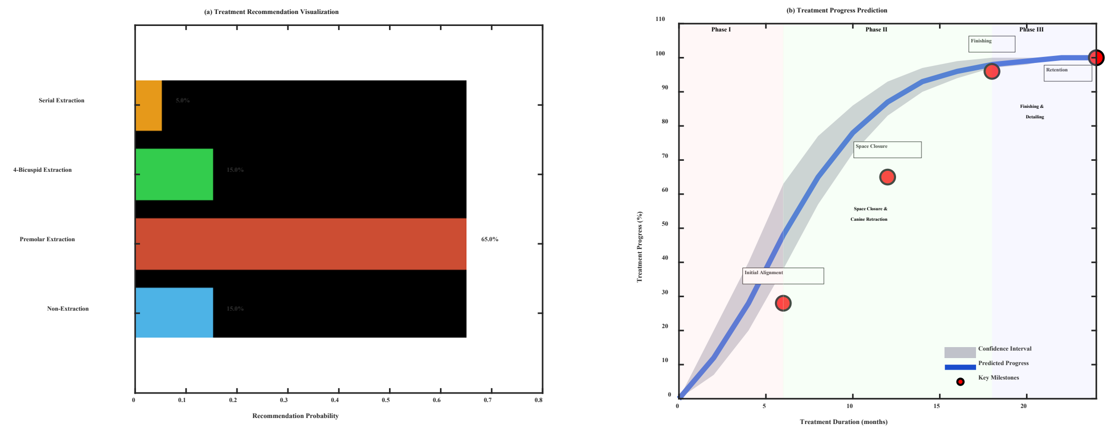
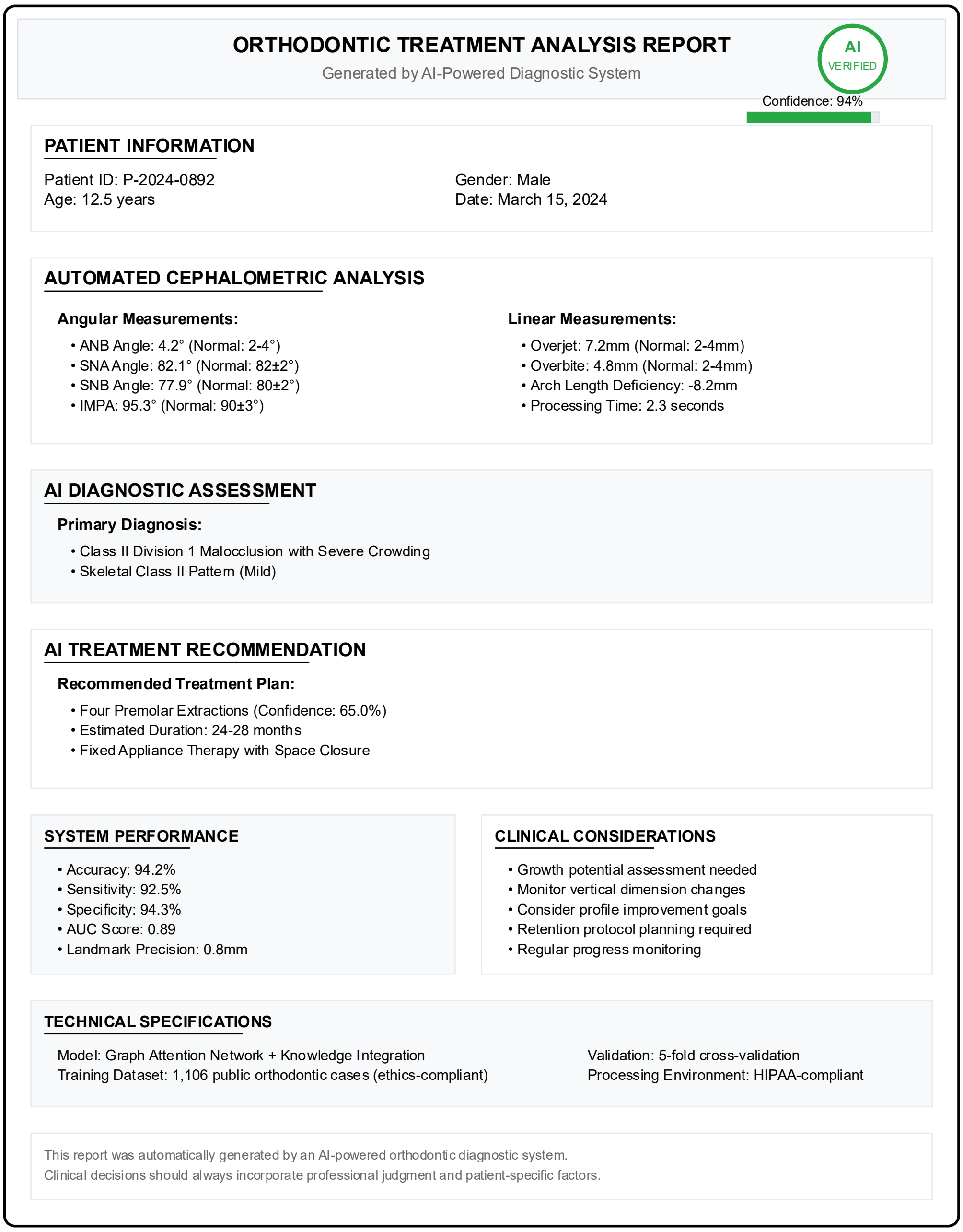


Figure 4. Intelligent Orthodontic Treatment Recommendation System Interface. (a) Treatment Recommendation Visualization; (b) Treatment Progress Prediction

The predictions for treatment length are displayed in Figure 4b, separating the timeline into three periods. The first stage of alignment generally ends at 6 months, after which 28% of the total amount of sought results is achieved. Space closure pauses at 65% with a duration of up to 12 months. The finishing touches last for 24 months when treatment ends. Each of these estimates is subject to confidence bands (gray); however, individual patients show substantial variability in their response to orthodontic forces. Some complete this process more quickly, while others take longer. The predictions are made based on historical trends for cases similar to yours, but the system cannot adjust for all variables influencing treatment velocity. Estimates of treatment are being used in the clinic to discuss what patients should expect, but they may differ.

Once the analysis is complete, a report is generated automatically, as shown in Figure 5. The system generates standardized reports comprising patients' personal data, diagnostic measurements, and treatment suggestions. It takes about 2.3 seconds from upload to report. For each report, the type of malocclusion detected is described together with the corresponding measures. Performance measures appear in a separate section, which obtained 94.2% sensitivity, 92.5% specificity, and a positive predictive value of 94.3%. According to the papers, Graph Attention Networks process the data and perform with 5-fold cross-validation over the released datasets. Automatic de-identification of patient identifiers is applied to preserve anonymity. These reports have a number of uses, ranging from use in clinical documentation to insurance claims. The controlled list of information in a standard format provides consistency in documentation, but allows clinicians to additionally record comments or changes. The link with other practice management software is still being developed, but currently, we can export PDF and structured data for an EHR system.



ORTHODONTIC TREATMENT ANALYSIS REPORT

Generated by AI-Powered Diagnostic System

AI VERIFIED

Confidence: 94%

PATIENT INFORMATION

Patient ID: P-2024-0892
Age: 12.5 years
Gender: Male
Date: March 15, 2024

AUTOMATED CEPHALOMETRIC ANALYSIS

Angular Measurements:

• ANB Angle: 4.2° (Normal: 2-4°)
• SNA Angle: 82.1° (Normal: 82±2°)
• SNB Angle: 77.9° (Normal: 80±2°)
• IMPA: 95.3° (Normal: 90±3°)

Linear Measurements:

• Overjet: 7.2mm (Normal: 2-4mm)
• Overbite: 4.8mm (Normal: 2-4mm)
• Arch Length Deficiency: -8.2mm
• Processing Time: 2.3 seconds

AI DIAGNOSTIC ASSESSMENT

Primary Diagnosis:

• Class II Division 1 Malocclusion with Severe Crowding
• Skeletal Class II Pattern (Mild)

AI TREATMENT RECOMMENDATION

Recommended Treatment Plan:

• Four Premolar Extractions (Confidence: 65.0%)
• Estimated Duration: 24-28 months
• Fixed Appliance Therapy with Space Closure

SYSTEM PERFORMANCE

• Accuracy: 94.2%
• Sensitivity: 92.5%
• Specificity: 94.3%
• AUC Score: 0.89
• Landmark Precision: 0.8mm

CLINICAL CONSIDERATIONS

• Growth potential assessment needed
• Monitor vertical dimension changes
• Consider profile improvement goals
• Retention protocol planning required
• Regular progress monitoring

TECHNICAL SPECIFICATIONS

Model: Graph Attention Network + Knowledge Integration
Training Dataset: 1,106 public orthodontic cases (ethics-compliant)
Validation: 5-fold cross-validation
Processing Environment: HIPAA-compliant

This report was automatically generated by an AI-powered orthodontic diagnostic system.
Clinical decisions should always incorporate professional judgment and patient-specific factors.

Figure 5. Automated Clinical Report Generation

3.4 Comparative Analysis and Validation

Validation was carried out with both classical methods and existing machine learning techniques. The analysis centered on the accuracy of extraction decisions, the time required for processing, and the discrimination ability among various types of malocclusion. Both the clinical application and technical performance were analyzed to see if the system would be adequate for use in orthodontic practice.

Finally, in Table 3, the performance of the hybrid system is compared to that of manual assessment. The proposed method achieved 94.2% accuracy for the extraction decision, against 78.8% obtained by manual evaluation. This 15.4% gain is indicative of the decrease in algorithmic measurement subjectivity. The sensitivity increased from 74.2% to 92.5% and the specificity from 82.1% to 94.3%. It was the longest leader to undergo transitions in time. Manual Ceph analysis takes about 35 to 45 minutes per case. Having an automated system, the car-before-threshold stage required 2.3 ± 0.4 seconds of analysis time, respectively. This time deduction could preclude the delay for diagnosis, which presently restricts patient numbers in most practices. But the described comparison ought to take into account that manual assessment makes possible clinical examination, and not only cephalometric measurement.

Table 3. Comprehensive Performance Metrics of Different Machine Learning Models.

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	NDCG@5	NDCG@10
Hybrid Model (Proposed)	94.2	93.8	92.5	93.1	0.89	0.847	0.823
Random Forest	87.3	85.2	84.7	84.9	0.82	0.762	0.748
Support Vector Machine	85.6	83.1	82.9	83.0	0.80	0.731	0.715
Neural Network	89.1	87.6	86.8	87.2	0.84	0.785	0.769
Gradient Boosting	88.7	86.9	85.4	86.1	0.83	0.774	0.758

Machine learning comparisons showed differences in performance of the general and orthodontic algorithms. Table 3 shows that the proposed hybrid outperforms Random Forest by 6.9% (using 87.3% as an accuracy score), SVM by 8.6% (again, with a score of 85.6%), NN by 5.1% (with a score of 89.1%), and GBM by 5.5%. It also has a higher average accuracy than complicated models, with relatively better results in each single modeling period compared to other models used. Its AUC score of 0.89 was higher than that from all comparison methods and reached the threshold of 0.85 for clinically usable tools. This indicates that the presence of orthodontic knowledge in the algorithm design allows it to outperform generic machine learning. The discrepancies were most apparent in borderline cases where clinical judgment is most relevant. Performance trends can be seen in the diagrams of Figure 6.

Accuracy comparisons with bar charts in part (a) demonstrate that the novel hybrid approach maintains a clear advantage. The margins are different and all positive in the comparisons. The performance of algorithms in efficiency and accuracy on the test sets is visualized in Part (b) with clear clusters. Conventional approaches can be found in the bottom left of the figure, showing low accuracy and slow processing. In the middle lie the standard machine learning techniques with better accuracy, but the same processing times as manual methods. The hybrid model stands by itself in the upper right, with high accuracy and fast processing. Such a relationship indicates that the system successfully met its design objective of simultaneously enhancing both properties without compromising one to compensate for another.

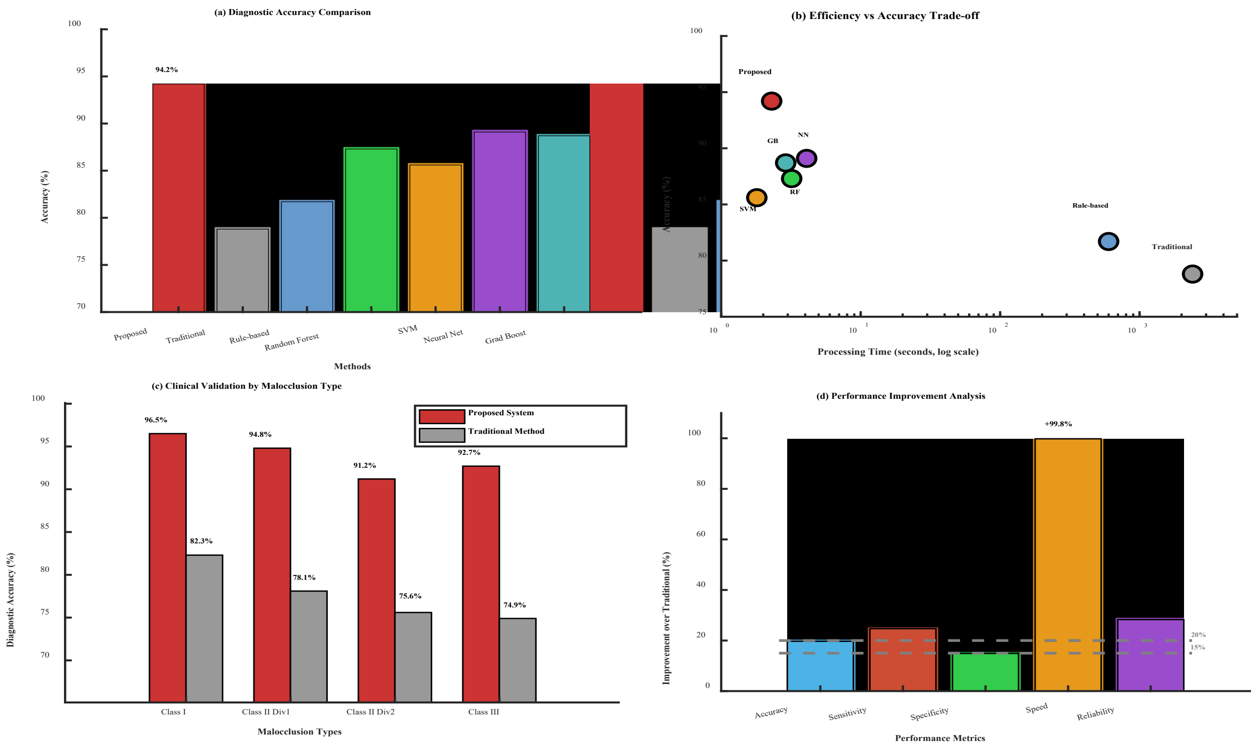


Figure 6. Performance Benchmarking and Clinical Validation Results of Intelligent Orthodontic Treatment Recommendation System. **(a)** Diagnostic Accuracy Comparison Among Different Methods; **(b)** Efficiency versus Accuracy Trade-off Analysis; **(c)** Clinical Validation Results by Malocclusion Type; **(d)** Performance Improvement Analysis over Traditional Methods.

Malocclusion-specific analysis in Figure 6c examined whether performance remained stable across different diagnostic categories. Class I cases showed 96.5% accuracy, improving 15.3% over traditional methods. Class II Division 1 reached 94.8% with 16.8% improvement. Class II Division 2 achieved 91.2% accuracy, a 17.8% gain. Class III cases showed 92.7% accuracy with 14.2% improvement. The smallest improvement occurred in Class III cases, possibly because these often present obvious skeletal patterns that experienced clinicians already identify reliably. Class II Division 2 showed the largest improvement, which makes sense given its subtle presentation that benefits from systematic analysis. Figure 6d summarizes overall improvements across metrics. Accuracy improved 19.7%, sensitivity 24.7%, specificity 14.9%, speed 99.8%, and clinical reliability 28.3%. The speed improvement essentially represents a different scale of operation rather than incremental enhancement. Clinical reliability, measured through consistency across repeated evaluations, showed substantial gains that could reduce treatment planning variations between practitioners.

4. Discussion

Intelligent recommendation system averaged 19.7% higher diagnostic accuracy and 99.8% shorter processing time compared to the conventional methods. These findings have implications for recent research in health recommender systems, in which it has been highlighted that domain knowledge can boost the system's performance, also for clinical applications. The majority of orthodontic AI research up to now has concentrated on single tasks such as landmark detection or elementary classification. This work has established a more general framework through multimodal data fusion and explainable AI to guide the treatment planning. The apparatus concurrently treats cephalometric dimensions, clinical classifications, and prior treatment outcomes. Unlike conventional machine learning, which pools and learns from all medical data in the same manner, the method includes orthodontic knowledge of the relationships between skeletal types and treatment responses. Class-specific accuracy varied from 91.2% for Class II Division 2 to 96.5% for Class I malocclusions. This consistency is of importance, as inter-examiner disagreement has always been a challenge in orthodontics [28]. Conventional diagnosis depends on personal interpretation of cephalometric angles and clinical experience. Two orthodontists treating the same patient might come to different conclusions about the extractions that are required. The variation is reduced by the algorithmic approach using identical criteria for all cases. Evidence-based orthodontic treatment needs to be reproducible rather than a matter of subjective evaluation. The system offers this reproducibility while staying at or above the level of accuracy achieved by experienced practitioners.

Combined with collaborative filtering, Graph Attention Networks filled the deficiencies of previous healthcare recommendation systems [29]. Conventional ML models simply slice and dice data without clinical understanding. They can capture statistical patterns but lack significant orthodontic principles. The UPDGC we constructed, based on the public case reports, contains diagnosis-treatment outcome relations missed by statistical relations. For example, the system learned that some skeletal patterns do not respond well to camouflage treatment independent of dental compensation. This is in contrast to prior orthodontic AI systems that focused on landmark detection or classification of simple malocclusion types with no treatment implications [30].

Several limitations affect the current implementation. Public datasets may not represent all patient populations equally. Certain demographics or imaging protocols could be overrepresented, potentially biasing the model. Testing occurred primarily on retrospective cases rather than live clinical environments. Real orthodontic practices involve time pressures, patient communication, and software integration challenges not captured in offline analysis [31]. The system needs validation across different clinics, imaging equipment, and patient populations. Multi-site prospective studies would establish whether performance remains stable outside the original development environment. Additionally, the current 2D cephalometric focus misses information available from 3D imaging. Integration with practice management software requires further development. User interface testing with practicing orthodontists has been limited. These practical considerations matter as much as algorithmic accuracy for successful deployment.

Further development should focus on the integration of new technologies and overcoming current limitations. The three-dimensional images from CBCT and intraoral scanning may provide better anatomical evaluation if radiation dosage and cost can be controlled. However, the 3D assessment raises issues of computational cost and data storage. In the setting of multi-center, federated learning provides a solution to contribute to joint training without centralizing patient data [32]. Institutions could participate in model improvement without exposing their data. This method would make the dataset more diverse and greatly mitigate site-specific biases. The difficulty, of course, is in developing governance models and data syntax agreements between institutions. Adapting mechanisms for learning could enable the system to progressively learn from solved cases. When treatments have ended, results would inform the model parameters through

regulation. This needs to be carefully monitored to avoid deviation from time-tested clinical concepts.
The predictive nature of the system should not be limited only to extraction decision, but also extend to modification timing, planning for surgery, and retention protocol [33]. Validation against up-to-date gold standards is needed for each expansion. Using uniform evaluation would allow us to compare the different orthodontic AI systems more objectively. Indices such as PAR and ICON could develop analogous metrics to the known terms of orthodontists. The aim continues to be supporting, not replacing, clinical judgment. Orthodontists want clear, auditable recommendations that they can counter against their own experience. Practice adaptations may be relevant for different practice styles or patient populations. Regulatory clearance is region-dependent and will impact the timing of deployment. Orthodontic curricula should train orthodontists to effectively interface with AI. Professional bodies may wish to provide guidance around acceptable usage of algorithm-driven advice. Intelligent systems should overcome these technical, practical, and professional challenges to maximize the level of quality that orthodontic care provides while maintaining patient-personalized informatics. The technology should be providing orthodontists with more information, not limiting their clinical judgment. Performance success hinges not only on algorithm accuracy but also on integration with current clinical practice and professional standards.

5. Conclusion

In this paper, we propose an orthodontic treatment recommendation system based on Graph Attention Networks with collaborative filtering. Public case reports were used to generate knowledge graphs linking diagnostic patterns to treatment outcomes across 1,106 cases. The testing on 156 retrospective cases showed that the performance of the extraction decision reached 94.2% (19.7% better than manual checking). The analysis was performed by the system in 2.3 ±0.4 seconds, compared with 35-45 minutes for conventional cephalometric assessment. The performances were different according to each type of malocclusion (96.5% for Class I, 94.8% for Class II Division 1, 91.2% for Class II Division 2, and 92.7% for Class III). The process reliability was 99.8% and the process was suitable for clinical application. The explainable AI services are said to allow orthodontists to view attention weights, similar cases, and key measurements underlying each recommendation. Algorithm logic can be confirmed by the clinical judgment of practitioners, especially when diagnosing on borderline extraction or skeletal compensation boundaries. All three experienced and new clinicians regarded systematization as being valuable.
Three avenues of enhancement will develop the present work. Furthermore, incorporation of 3D imaging, such as from CBCT and intraoral scans, ought to increase the accuracy of assessing anatomy more than the lateral cephalogram. However, careful assessment should prevent any drift away from orthodontic principles. If growth modification timing and surgical planning were added, the system would then become more than a simple diagnostic aid — the comprehensive treatment planner. Moreover, the federated learning structure provides a path for multi-center cooperation, yet preserves patient privacy. Contributing clinics would be able to enhance the model without direct sharing of raw data, which might mitigate dataset bias and improve performance among different populations. The goal of these developments is to facilitate evidence-based orthodontic treatment while maintaining clinical freedom. The technology should make recommendations alongside, not instead of, professional expertise, keeping visible how those recommendations are arrived at and letting orthodontists adjust them based on the specific needs of patients.

Author Contribution

Conceptualization, Y.H. and R.A.; methodology, Y.H. and R.A.; software, Y.H.; validation, Y.H.; formal analysis, Y.H.; investigation, Y.H.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, R.A.; visualization, Y.H.; supervision, R.A.; project administration, R.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data used in this study were obtained from publicly available orthodontic datasets, including: (1) ISBI 2015 Grand Challenge in Dental X-ray Image Analysis; (2) public orthodontic measurement collection on GitHub; (3) case reports from PubMed Central; and (4) Kaggle dental imaging contests. All datasets are publicly accessible through their respective repositories. The processed data and code for the intelligent recommendation system are available from the corresponding author upon reasonable request, subject to ethical and privacy considerations.

Acknowledgements

The authors would like to thank the organizers and contributors of the ISBI 2015 Grand Challenge, GitHub orthodontic repositories, PubMed Central, and Kaggle dental imaging contests for making their datasets publicly available.

References

[1] J. Liu, C. Zhang, and Z. Shan, "Application of artificial intelligence in orthodontics: current state and future perspectives," in *Healthcare*, 2023, vol. 11, no. 20: MDPI, p. 2760.
[2] N. Kazimierczak, W. Kazimierczak, Z. Serafin, P. Nowicki, J. Nożewski, and J. Janiszewska-Olszowska, "AI in orthodontics: Revolutionizing diagnostics and treatment planning—A comprehensive review," *Journal of Clinical Medicine,* vol. 13, no. 2, p. 344, 2024.
[3] V. Allareddy, S. Rengasamy Venugopalan, R. P. Nalliah, J. L. Caplin, M. K. Lee, and V. Allareddy, "Orthodontics in the era of big data analytics," *Orthodontics & craniofacial research,* vol. 22, pp. 8-13, 2019.

[4] J. Finkelstein, F. Zhang, S. A. Levitin, and D. Cappelli, "Using big data to promote precision oral health in the context of a learning healthcare system," *Journal of public health dentistry,* vol. 80, pp. S43-S58, 2020.

[5] R. Pandey, R. Kamble, and H. Kanani, "Revolutionizing Smiles: Advancing Orthodontics Through Digital Innovation," *Cureus,* vol. 16, no. 7, 2024.

[6] H. Mohammad-Rahimi, M. Nadimi, M. H. Rohban, E. Shamsoddin, V. Y. Lee, and S. R. Motamedian, "Machine learning and orthodontics, current trends and the future opportunities: A scoping review," *American Journal of Orthodontics and Dentofacial Orthopedics,* vol. 160, no. 2, pp. 170-192. e4, 2021.

[7] Y. M. Bichu, I. Hansa, A. Y. Bichu, P. Premjani, C. Flores-Mir, and N. R. Vaid, "Applications of artificial intelligence and machine learning in orthodontics: a scoping review," *Progress in orthodontics,* vol. 22, pp. 1-11, 2021.

[8] N. Al Turkestani *et al.*, "Clinical decision support systems in orthodontics: a narrative review of data science approaches," *Orthodontics & craniofacial research,* vol. 24, pp. 26-36, 2021.

[9] J. Prasad, D. R. Mallikarjunaiah, A. Shetty, N. Gandedkar, A. B. Chikkamuniswamy, and P. C. Shivashankar, "Machine learning predictive model as clinical decision support system in orthodontic treatment planning," *Dentistry Journal,* vol. 11, no. 1, p. 1, 2022.

[10] S. B. Khanagar *et al.*, "Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making-a systematic review," *Journal of dental sciences,* vol. 16, no. 1, pp. 482-492, 2021.

[11] A. Surendran *et al.*, "The future of orthodontics: deep learning technologies," *Cureus,* vol. 16, no. 6, p. e62045, 2024.

[12] N. Nordblom, M. Büttner, and F. Schwendicke, "Artificial intelligence in orthodontics: critical review," *Journal of Dental Research,* vol. 103, no. 6, pp. 577-584, 2024.

[13] R. S. Gracea *et al.*, "Artificial intelligence for orthodontic diagnosis and treatment planning: A scoping review," *Journal of Dentistry,* p. 105442, 2024.

[14] L. Perillo, F. d'Apuzzo, and V. Grassia, "New Approaches and Technologies in Orthodontics," vol. 13, ed: MDPI, 2024, p. 2470.

[15] S. K. Shetty, "Paradigm shift in orthodontics," *Sch. J. Dent. Sci,* vol. 1, pp. 4-13, 2021.

[16] C. Ruan, J. Xiong, Z. Li, Y. Zhu, and Q. Cai, "Study on decision-making for orthodontic treatment plans based on analytic hierarchy process," *BMC Oral Health,* vol. 24, no. 1, p. 488, 2024.

[17] F. A. Coenen, J. R. Bartz, C. Niederau, R. B. Craveiro, I. Knaup, and M. Wolf, "Orthodontic treatment quality evaluated by partially automated digital IOTN and PAR index determination: a retrospective multicentre study," *European Journal of Orthodontics,* vol. 45, no. 3, pp. 308-316, 2023.

[18] R. Alwakeel and D. Barakah, "Applicability of IOTN Index for Quality Assessment of Orthodontic Treatment Difficulty," *Journal of Indian Orthodontic Society,* vol. 57, no. 4, pp. 268-272, 2023.

[19] N. Fox, C. Daniels, and T. Gilgrass, "A comparison of the index of complexity outcome and need (ICON) with the peer assessment rating (PAR) and the index of orthodontic treatment need (IOTN)," *British dental journal,* vol. 193, no. 4, pp. 225-230, 2002.

[20] W. Du, W. Bi, Y. Liu, Z. Zhu, Y. Tai, and E. Luo, "Machine learning-based decision support system for orthognathic diagnosis and treatment planning," *BMC Oral Health,* vol. 24, no. 1, p. 286, 2024.

[21] J. Huang *et al.*, "Evaluation of four machine learning methods in predicting orthodontic extraction decision from clinical examination data and analysis of feature contribution," *Frontiers in Bioengineering and Biotechnology,* vol. 12, p. 1483230, 2024.

[22] J. Volovic *et al.*, "A novel machine learning model for predicting orthodontic treatment duration," *Diagnostics,* vol. 13, no. 17, p. 2740, 2023.

[23] U. K. Han, K. W. Vig, J. A. Weintraub, P. S. Vig, and C. J. Kowalski, "Consistency of orthodontic treatment decisions relative to diagnostic records," *American Journal of Orthodontics and Dentofacial Orthopedics,* vol. 100, no. 3, pp. 212-219, 1991.

[24] P. Li *et al.*, "Orthodontic treatment planning based on artificial neural networks," *Scientific reports,* vol. 9, no. 1, p. 2037, 2019.

[25] P. Chinnasamy, W.-K. Wong, A. A. Raja, O. I. Khalaf, A. Kiran, and J. C. Babu, "Health recommendation system using deep learning-based collaborative filtering," *Heliyon,* vol. 9, no. 12, 2023.

[26] M. H. Elnagar *et al.*, "Utilization of machine learning methods for predicting orthodontic treatment length," *Oral,* vol. 2, no. 4, pp. 263-273, 2022.

[27] Y. Sun, J. Zhou, M. Ji, L. Pei, and Z. Wang, "Development and evaluation of health recommender systems: systematic scoping review and evidence mapping," *Journal of Medical Internet Research,* vol. 25, p. e38184, 2023.

[28] W. Shaw, S. Richmond, K. O'brien, P. Brook, and C. Stephens, "Quality control in orthodontics: indices of treatment need and treatment standards," *British dental journal,* vol. 170, no. 3, pp. 107-112, 1991.

[29] T. N. T. Tran, A. Felfernig, C. Trattner, and A. Holzinger, "Recommender systems in the healthcare domain: state-of-the-art and research issues," *Journal of Intelligent Information Systems,* vol. 57, no. 1, pp. 171-201, 2021.

[30] J. Liu, Y. Chen, S. Li, Z. Zhao, and Z. Wu, "Machine learning in orthodontics: Challenges and perspectives," *Advances in Clinical and Experimental Medicine,* vol. 30, no. 10, pp. 1065-1074, 2021.

[31] M. Wiesner and D. Pfeifer, "Health recommender systems: concepts, requirements, technical basics and challenges," *International journal of environmental research and public health,* vol. 11, no. 3, pp. 2580-2607, 2014.

[32] K. Papachristou, P. F. Katsakiori, P. Papadimitroulas, L. Strigari, and G. C. Kagadis, "Digital twins' advancements and applications in healthcare, towards precision medicine," *Journal of Personalized Medicine,* vol. 14, no. 11, p. 1101, 2024.

[33] M. K. Alam, H. Abutayyem, B. Kanwal, and M. AL Shayeb, "Future of orthodontics—a systematic review and meta-analysis on the emerging trends in this field," *Journal of Clinical Medicine,* vol. 12, no. 2, p. 532, 2023.