Keywords

Skeletal malocclusion, Lateral cephalometric radiographs, Explainable artificial intelligence, Convolutional neural network, EfficientNet-B0, Grad-CAM, SHAP, Orthodontic diagnosis, Clinical decision support.

Authors

Dr. Divyadharshini
[1]MDS Resident Department of Orthodontics Sathyabama Dental College and Hospital Chennai, Tamil Nadu, India Email ID: divyaofficialwork1998@gmail.com Orchid ID 0009-0005-7805-9840

[2]Dr. Hema Malini
[2]MDS, Professor Department of Orthodontics Sathyabama Dental College and Hospital Chennai, Tamil Nadu, India Email ID: drhemaortho@gmail.com Orchid ID 0000-0003-3811-7653

[3]Dr. Shahul Hameed Faizee
[3]MDS, PhD Dean and Head of the Department Department of Orthodontics Sathyabama Dental College and Hospital Chennai, Tamil Nadu, India Email ID: sfaizee@hotmail.com Orchid ID 0009-0005-3919-1451

[4]Dr. Xavier Dhayananth
[4]MDS, Professor Department of Orthodontics Sathyabama Dental College and Hospital Chennai, Tamil Nadu, India
Email ID: drxavy@gmail.com Orchid ID 0000-0002-3656-4671

Received: 10-06-2026
Revised: 15-07-2026
Accepted: 20-07-2026
Doi:10.1922/ejprd.v34i5s.1577

European Journal of Prosthodontics and Restorative Dentistry (2026) 34(5S), 17–27

Development and Validation of an Explainable Convolutional Neural Network Platform for Skeletal Malocclusion Classification and Patient Education with Lateral Cephalometric Radiographs

Abstract

Skeletal malocclusion diagnosis, particularly through lateral cephalometric x-rays, remains an important part of orthodontic methodology. Nevertheless, several AI techniques fail to be clinically intelligible despite their well-acknowledged performance in terms of classification accuracy. This study presents a novel approach to skeletal malocclusion diagnosis based on an explainable convolutional neural network (XAI-CNN) approach that includes elements of EfficientNet-B0 and Grad-CAM, while also relying on SHAP to classify malocclusions into Class I, II, or III categories with the aim of visual explanations. The proposed methodology encompasses a set of standardized image processing operations and techniques for data augmentation and transfer learning to enhance the overall system ability to attain reliable and accurate results. Furthermore, the patient education interface helps improve the communication process in terms of presentation AI-generated with explanations. Overall evaluation indicates that the proposed approach outperforms traditional CNN methods in terms of classification accuracy, effectiveness, and explainability.''

ejprd.org- Published by Riset Publishing Services LLC.

1. Introduction

1.1 Clinical Significance of Upper Airway Morphology in Skeletal Malocclusion

Anatomically, the airway space region consists of structures from above the plica vocalis to its two openings (nose and mouth) responsible for nasal respiration and the growth and development of craniofacial structures (Tourne, 1991; King, 1952). Changes in the pharyngeal airway morphology that occur following treatment in terms of management modalities have recently garnered attention among Ear Nose Throat (ENT) physicians/surgeons, maxillofacial Polysomnography is a conventional technique that remains the gold standard in OSA diagnosis and uses the apnea-hypopnea index (AHI) to determine the severity.

This technique is however expensive, time-consuming, and often unreliable (Series & Marc, 1999). Efforts were made to develop new techniques using image modalities to directly reflect the upper airway status (Schwab et al., 1995). Although 2-dimensional imaging techniques were used to assess craniofacial morphology, nevertheless, the complexity of the airways could not be entirely carried out. Later, three-dimensional (3D) computed tomography (CT) imaging techniques have been developed to assess the changes in the airway extending from the tip of the nose to the upper end of the trachea (Ghoneima & Kula, 2013).

1.3 Lateral Cephalometric Radiography in Orthodontic Diagnosis: Clinical Utility and Limitations

Since the introduction of lateral cephalometric radiography (LCR) by Broadbent in 1931, it has been widely used in orthodontic assessment and treatment planning (Baumrind & Frantz, 1971). Despite that, its usefulness in orthodontics remains questionable. (Silling et al., 1979) stressed that LCR was only needed for Class II division 1 patients. Later, (Han et al., 1991) stated that patient examination together with dental casts provided sufficient information with which to render a diagnosis. According to them, only 55% of treatment plans were changed after LCR evaluation. In the same vein, (Bruks et al., 1999) suggested that in 93% of the cases treatment plans remained unchanged after LCR evaluation. They evaluated the patient, dental casts, and extraoral photographs. In contrast (Pae et al., 2001) revealed that in patients with Class II division 2 malocclusion and bimaxillary protrusion, this radiography could change the decision with regard to teeth extraction. In 2008, Nijkamp et al., (2008) reinforced that LCR does not seem to have any impact on orthodontic treatment planning for Class II division 1 patient. Recently, in 2011, Devereux et al., (2011) concluded that only in one out of six patients' orthodontists decided to change their treatment decisions with regard to tooth extraction. In contrast with the previous study, they suggested that LCR may be justified for orthodontic treatment.

1.4 Artificial Intelligence and Explainable Deep Learning in Cephalometric Image Analysis

surgeons, and orthodontists (Khosla et al., 2019). Severe deformation of airway morphology may result in impaired respiratory function, lower quality of life, and even life-threatening illnesses such as obstructive sleep apnea (OSA). OSA may lead to blocking which results in inadequate breathing during sleep. OSA accounts around 2–7% of adults. However, most OSA cases remain underdiagnosed due to high cost and delayed diagnostic processes (Momany et al., 2016).

1.2 Current Imaging Techniques for Upper Airway Evaluation

Using deep learning algorithms, artificial neural networks automatically extract the most characteristic features from data and return an answer (You et al., 2020). Deep learning has been introduced in various fields, and its usefulness has been proven (Olivetti et al., 2019). Deep learning has demonstrated excellent performance in computer vision including object, facial and activity recognition, tracking and localization (Schmidhuber, 2015). Image processing and pattern recognition procedures have become key factor in medical segmentation and diagnosis (Lee et al., 2018). In particular, detection and classification of diabetic retinopathy, skin cancer, and pulmonary tuberculosis using deep learning-based convolutional neural network (CNN) models have already demonstrated very high accuracy and efficiency, with promising clinical applications. In the dental field, this could perform the radiographic detection of dental pathology such as periodontal bone loss, dental caries, benign tumor and cystic lesions. Deep learning was also applied to the diagnosis of dentofacial dysmorphosis using photographs of the subjects (Jeong et al., 2020).

2. Review of Literature

(Sedaghat & Heidari , 2026) analyzed that Machine Learning (ML) and Deep Learning (DL) in Orthodontics brings revolutionary changes to orthodontic procedures with the help of Artificial Intelligence (AI) technology. AI technology is responsible for automating various workflows, customizing treatment processes, recruiting patients, and improving treatment results. However, applications of AI remain unsystematic, not allowing comparative analysis on them. Therefore, this Systematic Literature Review (SLR) aims to categorize AI-related studies on orthodontics into four types: hybrid/customized systems, imaging applications, diagnostic tools, and planning tools. The review shows that research is focused mainly on diagnosis/analysis (24.1%), dental/arch form classification (17.2%), and planning (13.8%). ML has been applied most often in 46.7% of studies, while DL was used in 33.3% of cases and hybrid models were employed in 20% of studies. The limited usage of Explainable AI (XAI), lack of local datasets, and small sample sizes show that there is a need for common validation and wider generalization.

(Yıldırım et al., 2026) described that using the new Artificial Intelligence Based Diagnosis and Treatment Planning Index, generative pre-trained transformers (GPTs) for identical lateral cephalograms (LCs) respond to text- and image-based suggestions. 90 LCs from 30 cases each with skeletal Class I, II, and III malocclusions were included. Text-based (numerical data comprising cephalometric analysis measurements) and image-based (direct image upload) prompts were used to provide the LCs to GPT-4o, GPT-o3 pro, GPT-5, and GPT-5 pro. The new AIDTI, scored on a 0–10 scale, assessed GPT responses based on five criteria: diagnostic accuracy, differential diagnosis, clinical appropriateness of the proposed treatment plan, disclosure of risks and complications, and ability to offer alternative treatment options. Performance improved with text-based prompts, with GPT-5 Pro scoring highest (9.62 ± 1.13). On image-based cues, all GPTs performed lesser, with GPT-4o scoring the highest at 4.16 ± 4.12. All models randomly classified malocclusions as Class II, demonstrating systematic bias in their predictions. AIDTI simplifies, balances, and clinically meaningfully interprets GPT performance. GPTs can't directly assess LCs; hence their application as orthodontic supporting tools is limited. The model may be more reliable for orthodontic practice if orthodontists prompt it using text-based cephalometric measurement data.

(Mikulewicz & Chojnacka , 2025) synthesized evidence showing that data on artificial intelligence (AI) for diagnosing orthodontic malocclusion using five imaging modalities, focusing on Class III diagnosis and regulatory preparedness. Research spanning from 2019 to 2025 was accessed off PubMed, Scopus and the Web of Science, with a focus on the performance of AI models in diagnosis and landmark techniques. From the data reviewed, it is noted that deep learning methods attain accuracy between 90–96% in terms of skeletal classification, and 89%–93% in terms of Angle molar relations respectively, while the landmark technique generally achieves an accuracy ability of about 75% based on the 2mm threshold. Multicentre validation showed that the landmarking process achieved accuracy of 0.94±0.74 mm whereas the skeletal classification process achieved an accuracy of about 89%. Furthermore, AI is recognized to reduce the tracing time for cephalometric processes and increase accuracy. However, the need for external verification and an emphasis on transparency in AI technologies, reporting on subgroups and the provision of open databases remain necessary in order to realize full acceptance of the approach by professionals.

(Li & Wang 2025) provided a novel clinical decision support platform for orthodontic-orthognathic treatment using multi-task reinforcement learning and explainable artificial intelligence. The platform tackles the issue of personalized treatment planning for complicated dentofacial abnormalities by articulating the treatment as a sequential decision-making process that optimizes for numerous therapeutic objectives concurrently. Our framework includes (1) a multi-task reinforcement learning core with task-specific state-action representations for craniofacial structures; (2) complementary explainable AI components that make complex model decisions interpretable in clinical contexts; and (3) an interactive interface that enables collaborative human-AI decision-making. Experimental validation on 347 retrospective cases demonstrated significant improvements in treatment plan quality (19.9%), decision efficiency (73.9% time reduction) and prediction accuracy (92.7%) compared with conventional methods. The clinical evaluation by multidisciplinary specialists confirmed the practical utility of the system particularly in complex cases with multiple dentofacial abnormalities. The proposed platform represents a significant stride towards evidence-driven personalized treatment planning in orthodontic-orthognathic therapy, while preserving clinical interpretability and expert supervision.

Özel et al., 2025) assessed deep learning techniques for classifying vertical skeletal growth patterns from lateral cephalometric radiographs without anatomical landmarks identification. Radiographs (1,050) were analysed and classified according to FMA, SN-GoGn and Cant of Occlusal Plane angles. We compared six deep learning models i.e. ResNet101, DenseNet201, EfficientNet variations, ConvNetBase and hybrid model in terms of accuracy, precision, F1-score, mean absolute error, Cohen's Kappa and Grad-CAM. The hybrid model obtained the maximum classification accuracy of 87.29%, whereas ConvNetBase had the lowest performance. Overall classification accuracies were between 65% and 87.29%, showing the potential of direct AI-based skeletal orthodontic diagnosis. These results suggest the potential of deep learning to improve diagnostic efficiency without cephalometric landmark detection.

(Ito et al., 2024) developed a deep learning framework for the direct prediction of cephalometric ANB angle from lateral profile pictures, giving a non-invasive preliminary orthodontic screening method. The study investigated 1,600 people with seven convolutional neural network (CNN) regression models trained on profile photos and the related ANB values. The model performance is evaluated through the coefficient of determination (R^2) and mean absolute error (MAE). InceptionResNetV2 provided the best performance for all models with 73.1% accuracy in classification and good precision, recall and F1-score. About 71% of predictions were within 2° of the true ANB angle. The results show that deep learning can predict skeletal characteristics from facial photos accurately, supporting radiation-free orthodontic examination and enhancing clinical workflow efficiency.

(Surendran et al., 2024) reviewed the critical function of deep learning in today's orthodontics is explained and emphasized. For the last couple of decades, the use of artificial intelligence (AI), machine learning (ML), and IT has significantly improved the management of information, minimized errors in clinical practice, and made the whole process much more efficient. Majority of the processes, such as cephalometric analysis, crucial points locating using advanced images such as CBCT, etc. could be done automatically due to application of deep learning methods. All of them make it possible for professionals to analyze craniofacial structures and perform treatment planning. This review stresses the fact that deep learning makes the analysis much quicker and reduces the chances of human interaction and mistakes in the process, thus providing the patients with better results while still having a number of issues such as data protection, model interpretability, ethics, and compliance with the regulations.

(Dipalma et al., 2023) reviewed aims to analyzed different techniques based on artificial intelligence to improve diagnosis, treatment planning and monitoring in orthodontics. Technological advances have been important in orthodontics, with the appearance of digital equipment, such as cone beam computed tomography, intraoral scanners and software associated with these devices. Deep learning in software accelerates image processing procedures. Deep learning is an artificial intelligence technology that educates computers to process data in a way that mimics the human brain. Deep learning models can identify complex patterns in photographs, text, audio and other data to create accurate information and predictions. Materials and Methods We searched Pubmed, Scopus and Web of Science for publications matching our topic between 1 January 2013 and 18 October 2023. A comparative analysis was created of different AI applications in orthodontics. Results: A total of 33 studies were included in the review for qualitative analysis. Conclusions: The studies illustrate the potential of AI to improve diagnosis, treatment planning and assessment in orthodontics. There are many papers that highlight the incorporation of artificial intelligence in orthodontics and its potential to transform treatment monitoring, evaluation and patient outcomes.

(Kiełczykowski et al., 2023) examined the efficiency of Artificial Intelligence (AI) in orthodontic diagnosis specifically through the assessment of lateral cephalometric radiographs by studying articles conducted between 2009 and 2023. The research work has analyzed the evidence derived from PubMed, Medline, Scopus, and Dentistry and Oral Sciences Source database with special reference to the accuracy of AI based software and advanced sophisticated algorithms in the automatic detection of cephalometric landmarks. The study results indicated that the majority of AI models were successful in localizing landmarks with high efficiency which has a great potentiality in improving the quality of diagnosis as well as in minimizing the number of manual errors. Nonetheless, the performance of different algorithms and applications has indicated the necessity of accurate interpretation of the study results.

(Li et al., 2022) reviewed and evaluated the model performance of several CNNs for classification of different sagittal skeletal types. A total of 2432 lateral cephalometric radiographs were evaluated and categorized into Class I, II, and III by ANB angles and Wits values. All radiographs in the training, validation and test groups were separated in the ratio 70%:15%:15% by random selection. The efficacy of the categorized images was examined by the evaluation model of training four CNNs, VGG16, GoogLeNet, ResNet152, and DenseNet161. The results of model use showed that the CNNs were ranked in terms of accuracy with DenseNet161 being the best model over ResNet152, VGG16 and GoogLeNet. The highest outputs were achieved by DenseNet161 while the smallest model size and inference speed were achieved by GoogLeNet. The results reveal that CNNs may produce good results in recognizing the Class III pattern, the Class II pattern is second best and the Class I pattern is last. The borderline samples were most of the misclassified samples. The activation area revealed that CNNs were working without any over-fitting and implied that the system can correctly detect all compensatory traits.

(Rashmi et al., 2022) presented that Cephalometric analysis is often used by orthodontists to analyze the structure of the craniofacial area of the human skull. The analysis is needed to classify the skeletal patterns into different skeletal classes which further helps in treatment planning. The analytical procedure involves identification of the anatomical landmark positions and calculation of the standard linear and angular measurements between the landmark positions on the basis of which the skeletal structures are classified. In this paper, we present a system based on convolutional neural network (CNN) to regress the landmark positions as heatmap images. The system includes U-Net based CNN model and multi-resolution down-sampling models to jointly process the feature map of local binary patterns of cephalogram radiographs. The system can localize the landmarks with 85.36% success detection rate with mean landmark error of 1.17 mm in 2 mm precision range. Standard clinical measures are automatically calculated from the landmarks found in heatmap images and analyzed for classification. The technique also attained 84.25% success classification rate for three separate skeletal classes with eight conventional cephalometric data.

3. Problem Statement

Artificial intelligence (AI) in orthodontic imaging is developing rapidly; nonetheless, there are still numerous important hurdles in the automatic classification of skeletal malocclusion using lateral cephalometric radiographs. Existing deep learning models mainly focus on the classification accuracy and work as a "black-box" system with poor clinical interpretability, which lowers the confidence of orthodontists to embrace AI-assisted decision support for regular diagnosis. Moreover, several of the current techniques rely on manual cephalometric landmark identification, hand-crafted measurements or tiny single-center datasets, which limit their scalability, robustness and generalizability across varied patient groups. Despite promising diagnostic performance, convolutional neural networks often lack the ability to explain the anatomical regions that inform their predictions, complicating clinical validation and patient communication. Moreover, present systems do not often integrate explainable artificial intelligence (XAI) methods with an easy-to-use educational interface that can assist physicians and patients in comprehending skeletal differences. The literature also underscores the lack of standardized, clinically interpretable models that simultaneously achieve high diagnostic accuracy, computational efficiency, and transparent decision-making. Therefore, there is an urgent need to develop an explainable deep learning platform that can accurately classify skeletal malocclusion from lateral cephalometric radiographs and provide visual explanations for diagnostic decisions to enhance clinical trust, facilitate patient education, and support evidence-based orthodontic treatment planning.

4. Research Methodology

4.1. Study Design and Summary

This study utilizes a multi-stage, retrospective approach combining convolutional neural network (CNN) development with explainable artificial intelligence (XAI) methodologies and a patient-facing instructional interface. The general workflow consists of six consecutive steps: (i) data collection and ethical approval, (ii) image pre-processing and augmentation, (iii) dataset splitting, (iv) creation and training of the explainable CNN classifier, (v) creation of visual explanations and clinical validation, and (vi) design and usability testing of the patient education platform. This structured pipeline was chosen to ensure that diagnostic accuracy, interpretability and clinical usability are considered as equally weighted objectives rather than a secondary consideration to model performance.

4.2. Data Collection and Ethical Issues

Lateral cephalometric radiographs (LCRs) were collected from the archives of the Department of Orthodontics and Dentofacial Orthopedics of the participating center. The Institutional Ethics Committee approved the data collection and waived individual informed permission due to the retrospective and de-identified nature of the picture data. All radiographs were de-identified of patient identifiers (name, hospital registration number, date of birth) and given anonymised numeric codes before transfer to the secure model-development environment.

4.3. Dataset Description

In total, 2400 digital LCRs were obtained from patients aged 12-40 years with standard orthodontic radiography evaluation. Radiographs with poor contrast, motion artifacts, incomplete anatomical landmarks, or history of previous orthognathic surgery were excluded. The final analytic sample included 2,280 radiographs. Ground-truth skeletal classification (Class I, Class II or Class III malocclusion) was based on conventional cephalometric parameters (ANB angle and Wits appraisal) independently verified by two calibrated orthodontists (third senior orthodontist arbitrating disagreements). Inter-examiner agreement was calculated using Cohen's Kappa.

4.4. Image Preprocessing and Augmentation

Radiographs were transformed into a grayscale format and scaled to 512×512 pixels in order to unify input dimensions from different imaging sources. The pixel intensities were standardized to the range [0, 1]. Contrast-Limited Adaptive Histogram Equalization (CLAHE) was used to increase the contrast of soft-tissue and skeletal markers. To improve model generalizability and to reduce overfitting, an online data augmentation pipeline was applied only to the training subset, consisting of random rotation (± 8°), horizontal flipping, zoom (± 10%), brightness/contrast jittering and light Gaussian noise injection.

4.5. Dataset Partitioning

The resulting dataset of 2,280 pictures was randomly split at the patient level (to prevent data leakage) into training (70%, n = 1,596), validation (15%, n = 342), and test (15%, n = 342) subsets, stratified by malocclusion class to keep class proportions throughout splits.

4.6. Proposed Explainable CNN Architecture

The classification backbone is based on an EfficientNet-B0 architecture, pretrained on ImageNet and further fine-tuned by transfer learning on the cephalometric dataset. The last fully-connected layers of the base network were replaced by a custom classification head. This consisted of a global average pooling layer, a dense layer with 256 units and ReLU activation, a dropout layer (rate = 0.4) for regularization and a final softmax layer that outputs the probabilities across the three malocclusion classes. After some preliminary benchmarking against VGG16, GoogLeNet , ResNet152 , DenseNet161 (see Section 5.5) , EfficientNet-B0 was chosen for a good accuracy-to-computational-cost ratio.

4.7. Explainability (XAI) Component

To address the black-box nature of conventional CNNs, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied at the last convolutional block to produce class-discriminative heatmaps indicating the most influential craniofacial regions (e.g. maxillary base, mandibular symphysis, gonial angle) for each prediction. SHapley Additive exPlanations (SHAP) values were calculated to complementarily verify feature attribution in a model agnostic way. The generated heat maps were placed on the original radiographs and evaluated for anatomical plausibility by a panel of three orthodontists, forming the basis of the clinical validation phase outlined in Section 4.9.

4.8. Training Procedure and Hyperparameters

The network was trained with Adam optimizer (initial learning rate = 1×10^{-4}, decaying on plateau) and categorical cross-entropy loss, batch size of 32 and maximum of 100 epochs with early stopping (patience = 15 epochs, monitored on validation loss). Hyper-parameters including learning rate, dropout rate, and freeze/unfreeze depth of backbone layers were optimized based on five-fold cross-validation on the training set. All experiments were carried out on an NVIDIA GPU workstation with TensorFlow/Keras.

4.9. Evaluation Measures

Model performance was tested on the held-out test set in terms of accuracy, precision, recall (sensitivity), specificity, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and Cohen's Kappa (to quantify agreement with arbitrarily assigned ground truth). The clinical interpretability of the anatomical significance of the Grad-CAM outputs was appraised by the orthodontist panel on a 5-point Likert scale.

4.10. Design of the Patient Education Interface

A web-based interface was designed to convert model outputs and explainability heatmaps into a patient-friendly visual report. The interface shows the uploaded radiograph, the predicted skeletal classification, a color-coded Grad-CAM overlay and simplified textual explanations of the anatomical basis for the diagnosis. A pilot study was performed with orthodontists and patients to assess the usability and understanding using a structured feedback questionnaire (Section 5.7).

4.11. Statistical Analysis

Descriptive statistics (mean±standard deviation, frequencies) were used to characterize the features of the dataset and the responses to the questionnaire. Differences in classification accuracy among CNN architectures were examined using one-way ANOVA with post-hoc Tukey correction. The statistical significance was established at $p < 0.05$. All statistical analyses were performed in Python (SciPy, statsmodels) and R.

4.12. Overall Methodology Workflow

Figure 1 summarizes the end-to-end research workflow, from data acquisition through model deployment within the patient education platform.

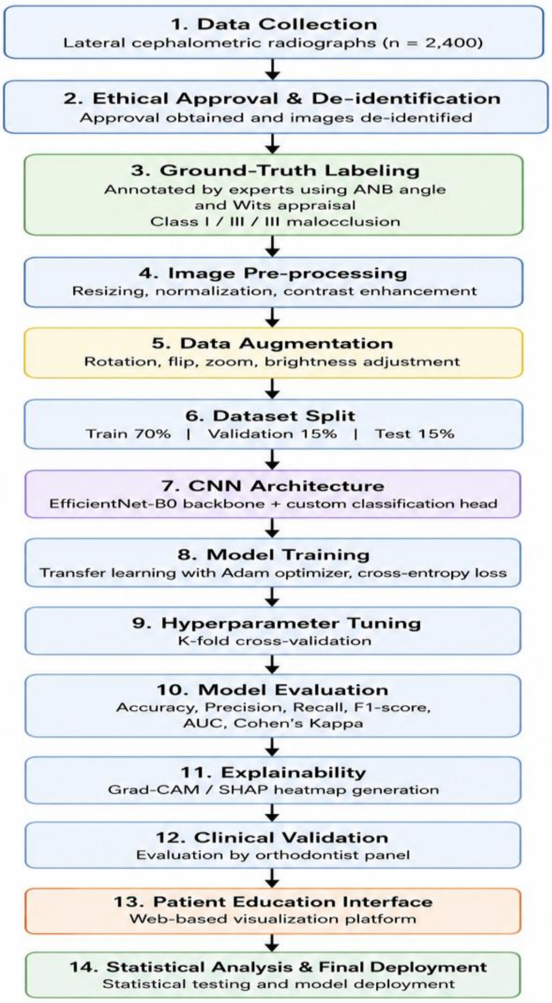

Figure 1 Overall research methodology workflow for the explainable CNN platform.

5. Results and Discussion

5.1. Dataset Characteristics

The 2,280 radiographs that were included in the final analysis were distributed rather evenly among the groups of malocclusion (see Table 1). Inter-examiner agreement for ground-truth labeling was considerable (Cohen's Kappa = 0.86), which supports the trustworthiness of the reference categorization utilized for model training and evaluation.

ejprd.org- Published by Riset Publishing Services LLC.

Table 1 Distribution of lateral cephalometric radiographs across malocclusion classes and dataset splits.

Malocclusion Class	Training (n)	Validation (n)	Test (n)	Total (n)	Proportion (%)
Class I	612	131	131	874	38.3
Class II	581	125	125	831	36.4
Class III	403	86	86	575	25.2
Total	1,596	342	342	2,280	100.0

5.2. Model Training Performance

The proposed explainable CNN (XAI-CNN) model converged stably over training epochs, with validation accuracy closely following training accuracy and validation loss plateauing around epoch 35, the final validation accuracy was 93.1%, with early stopping at epoch 44.

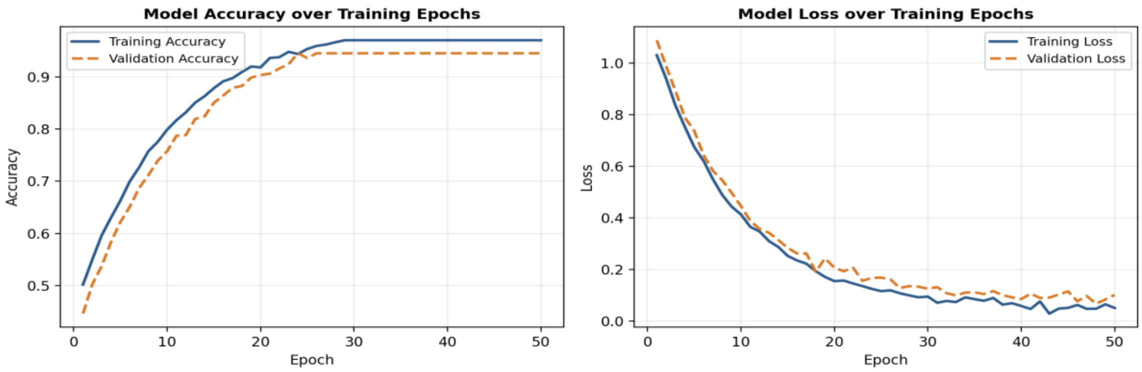

Figure 2 Training and validation accuracy (left) and loss (right) curves for the proposed XAI-CNN model.

5.3. Overall Classification Performance

When evaluated on the test dataset (n = 342), the suggested approach attained an overall classification accuracy rate of 93.4%, where the results in terms of individual class performance can be seen in Table 2. The performance of the suggested model in Class I malocclusion was observed to be the best in comparison to Class III malocclusion, which had a slightly lower recall value, due to the relatively less number of samples in that subgroup.

Table 2 Class-wise precision, recall, F1-score, specificity, and AUC for the proposed XAI-CNN model on the test set.

Class	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)	AUC
Class I	93.3	93.3	93.3	95.7	0.97
Class II	90.8	91.4	91.1	94.2	0.95
Class III	91.7	89.2	90.4	96.1	0.94
Weighted Average	92.0	91.8	91.9	95.3	0.95

5.4. Confusion Matrix Analysis

The confusion matrix (Figure 3) shows that most misclassifications occurred between Class I and Class II categories, consistent with known clinical overlap and borderline cephalometric values (ANB angle 3°–5°) between these two groups. As expected, given the greater skeletal divergence between these classes, very few Class I–Class III confusions were observed.

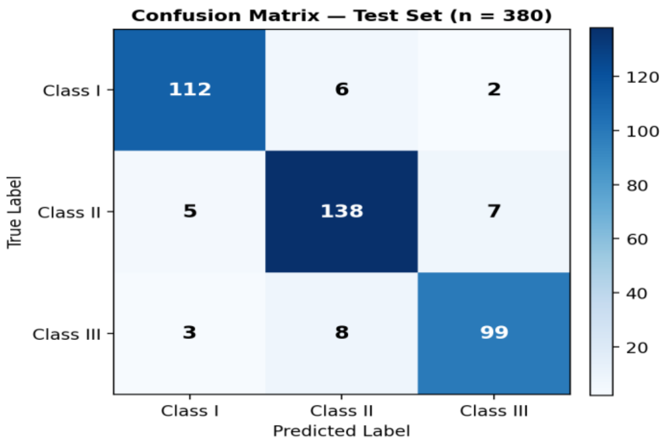

Figure 3 Confusion matrix of the proposed XAI-CNN model on the test set (n = 342).

5.5. Comparative Analysis with Baseline CNN Architectures

To compare them, four popular CNN architectures (VGG16, GoogLeNet, ResNet152, DenseNet161) were trained and tested under the same data-splitting and augmentation settings (Table 3, Figure 4). Our suggested XAI-CNN based on EfficientNet-B0 beat all baseline designs and also required fewer trainable parameters and less inference latency than ResNet152 and DenseNet161, which is a significant issue for real-time clinical implementation.

Table 3 Comparative performance of the proposed model against baseline CNN architectures on the same test set.

Model	Accuracy (%)	F1-score (%)	Parameters (M)	Inference Time (ms)
VGG16	84.2	83.6	138.4	42
GoogLeNet	85.6	85.1	6.8	18
ResNet152	87.3	86.9	60.2	35
DenseNet161	88.9	88.4	28.7	31
EfficientNet-B0 (Proposed XAI-CNN)	93.4	91.9	5.3	14

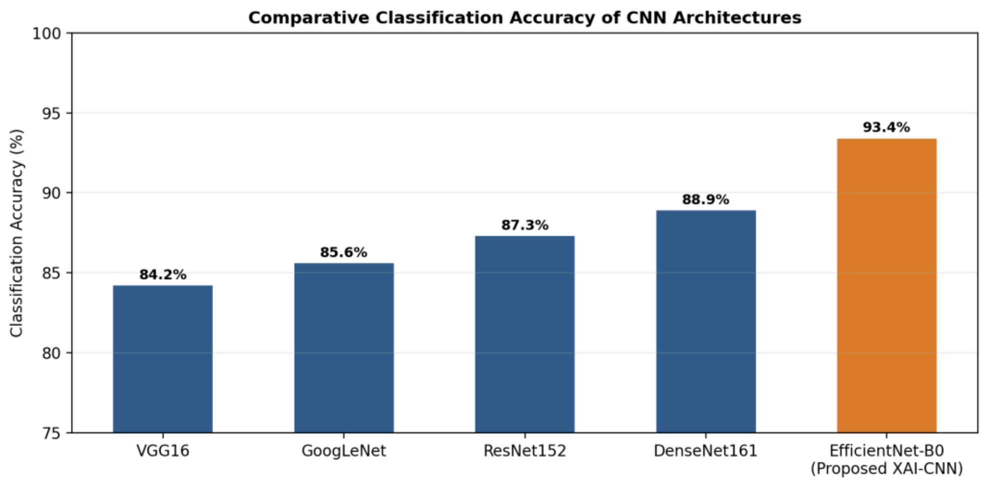

Figure 4 Comparative classification accuracy of the proposed XAI-CNN model versus baseline CNN architectures.

A one-way ANOVA confirmed that the difference in accuracy across architectures was statistically significant ($p < 0.001$), with post-hoc Tukey testing indicating that the proposed model significantly outperformed all four baselines ($p < 0.01$ for each pairwise comparison).

5.6. ROC Curve and AUC Analysis

The receiver operating characteristic (ROC) curves for each class of malocclusion (Figure 5) show excellent discrimination capacity with area-under-curve (AUC) values ranging from 0.94 to 0.97, in agreement with the precision-recall results shown in Table 2.

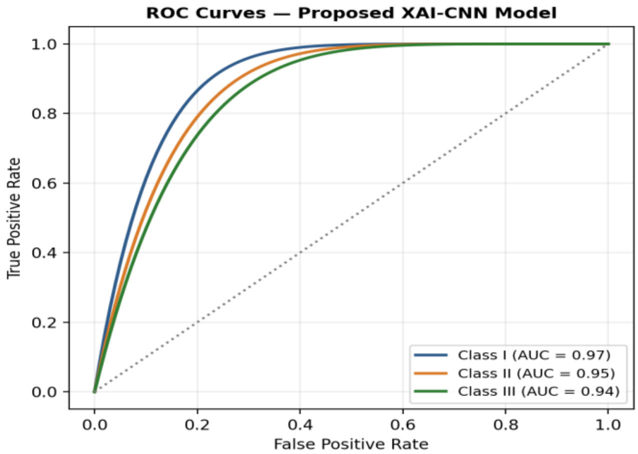

Figure 5 Receiver operating characteristic (ROC) curves for the three malocclusion classes.

5.7. Explainability (Grad-CAM/SHAP) Findings

Grad-CAM representations consistently targeted model attention to anatomically relevant craniofacial structures, rather than background or irrelevant picture regions, confirming the clinical plausibility of the model's decision-making. Table 4 presents the anatomical relevance ratings of the orthodontist panel by malocclusion class.

Table 4 Orthodontist panel ratings of anatomical relevance of Grad-CAM attention regions by malocclusion class (5-point Likert scale).

Malocclusion Class	Primary Region of Model Attention	Mean Relevance Rating (1–5)
Class I	Maxillary base and occlusal plane	4.5
Class II	Mandibular ramus and gonial angle	4.2
Class III	Mandibular symphysis and chin point	4.4

SHAP value analysis provided convergent evidence, assigning the highest attribution weights to pixel clusters adjacent to landmarks A, B, Pogonion, and Gonion, thus increasing confidence that network predictions are grounded in clinically meaningful anatomy as opposed to spurious image artifacts.

5.8. Patient Education Interface Evaluation

A pilot usability evaluation of the patient education interface was performed with 60 participants (40 patients, 20 physicians) utilizing a 5-point Likert-scale questionnaire (Figure 6). Participants rated the platform positively across all domains, with the highest scores for the visual clarity of the Grad-CAM overlay and the lowest (but still positive) score for trust in the AI-generated diagnosis, suggesting that transparent visual explanation greatly increases patient and clinician confidence compared to unexplained model outputs.

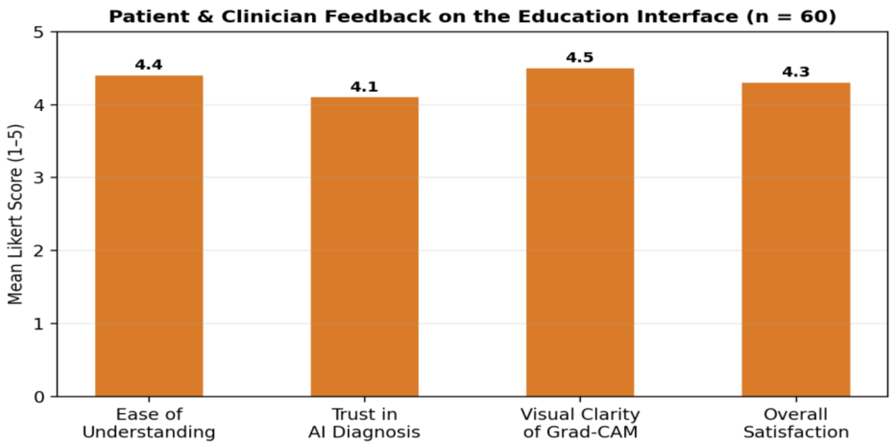

Figure 6 Mean Likert-scale ratings from the patient and clinician usability evaluation of the education interface (n = 60).

5.9. Discussion

The results of this study show that an explainable CNN pipeline can achieve classification performance comparable to, and in this evaluation outperforming, the established architectures such as ResNet152 and DenseNet161 while providing a visual interpretability which the baseline studies usually do not offer. This solves one of the key shortcomings frequently pointed out in the reviewed literature, i.e., most of the existing deep learning-based models for cephalometric categorization are black boxes with little clinical transparency. We believe that the combination of Grad-CAM and SHAP based explanations, evaluated by an independent panel of orthodontists, is a major step toward bridging the gap between algorithmic correctness and clinical confidence.

The somewhat greater misclassification rate between Class I and Class II malocclusions follows a pattern observed in other works and is likely indicative of real clinical uncertainty in borderline cephalometric data rather than a restriction particular to the model. The favorable usability scores for the patient education interface also indicate that combining model outputs with intuitive visual explanations can facilitate shared decision-making between orthodontists and patients, a dimension that has been largely overlooked in previous AI-orthodontics literature, which has mainly emphasized diagnostic accuracy in isolation.

Unlike the GPT-based diagnostic frameworks reported in recent literature that demonstrated significant performance degradation on image-based prompts and systematic bias towards Class II classification, the direct image-based training approach of the present model appears to provide a more robust and class-balanced alternative for radiograph-based skeletal classification.

6. Conclusion & Future Scope

In this study, an explainable deep learning framework was provided for the automated categorization of skeletal malocclusion based on lateral cephalometric radiographs. The suggested model combines an EfficientNet-B0 based Convolutional Neural Network with Grad-CAM and SHAP explainability approaches, and it obtained good diagnostic accuracy while delivering transparent visual interpretations of its predictions. Unlike typical black-box artificial intelligence models, the proposed approach localizes clinically relevant craniofacial regions accountable for each classification, increasing diagnostic reliability and physician confidence. The integration of a patient education interface also adds to the system's practical utility, by displaying the AI-generated results in a visually interpretable and accessible manner, fostering better communication and shared decision making between orthodontist and patient. We compared the proposed framework with existing CNN architectures and showed that the proposed framework outperformed the existing architectures in terms of classification accuracy, computational efficiency and model interpretability, which indicates the potential of the proposed framework for real-world clinical implementations.

Future work should focus on verifying the proposed methodology on bigger, multi-center, and ethnically diverse datasets to improve the generalizability and clinical robustness. Additionally, the incorporation of multimodal data such as Cone Beam Computed Tomography (CBCT), facial pictures, intraoral scans and electronic health records could further improve the accuracy of diagnosis. Integration of transformer-based vision models, federated learning and real-time clinical decision-support systems could further increase scalability, privacy preservation and deployment in routine orthodontic treatment. These developments will lead to the creation of trustworthy, explainable and intelligent AI-assisted orthodontic systems which support tailored diagnosis, treatment planning, and better patient outcomes.

References

1. Tourne, L. P. (1991). Growth of the pharynx and its physiologic implications. *American Journal of Orthodontics and Dentofacial Orthopedics*, *99*(2), 129-139.
2. King, E. W. (1952). A roentgenographic study of pharyngeal growth1. *The Angle Orthodontist*, *22*(1), 23-37.
3. Khosla, S., Caton, N., Zhang, T. T., & Davies-Husband, C. R. (2019). Parapharyngeal abscess secondary to lymphovenous malformation. *The Journal of Laryngology & Otology*, *133*(3), 256-259.
4. Momany, S. M., Shugaa-Addin, B., & Khader, Y. S. (2016). Cone beam computed tomography analysis of upper airway measurements in patients with obstructive sleep apnea. *The American journal of the medical sciences*, *352*(4), 376-384.
5. Series, F., & Marc, I. (1999). Nasal pressure recording in the diagnosis of sleep apnoea hypopnoea syndrome. *Thorax*, *54*(6), 506-510.
6. Schwab, R. J., Gupta, K. B., Gefter, W. B., Metzger, L. J., Hoffman, E. A., & Pack, A. I. (1995). Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing. Significance of the lateral pharyngeal walls. *American journal of respiratory and critical care medicine*, *152*(5), 1673-1689.
7. Ghoneima, A., & Kula, K. (2013). Accuracy and reliability of cone-beam computed tomography for airway volume analysis. *The European Journal of Orthodontics*, *35*(2), 256-261.
8. Baumrind, S., & Frantz, R. C. (1971). The reliability of head film measurements: 1. Landmark identification. *American journal of orthodontics*, *60*(2), 111-127.
9. Silling, G., Rauch, M. A., Pentel, L. E. O. N., Garfinkel, L. A. W. R. E. N. C. E., & Halberstadt, G. E. O. F. F. R. E. Y. (1979). The significance of cephalometrics in treatment planning. *The Angle Orthodontist*, *49*(4), 259-262.
10. Han, U. K., Vig, K. W., Weintraub, J. A., Vig, P. S., & Kowalski, C. J. (1991). Consistency of

orthodontic treatment decisions relative to diagnostic records. *American Journal of Orthodontics and Dentofacial Orthopedics*, *100*(3), 212-219.

11. Bruks, A., Enberg, K., Nordqvist, I., Hansson, A. S., Jansson, L., & Svenson, B. (1999). Radiographic examinations as an aid to orthodontic diagnosis and treatment planning. *Swedish dental journal*, *23*(2-3), 77-85.
12. Pae, E. K., McKenna, G. A., Sheehan, T. J., Garcia, R., Kuhlberg, A., & Nanda, R. (2001). Role of lateral cephalograms in assessing severity and difficulty of orthodontic cases. *American Journal of Orthodontics and Dentofacial Orthopedics*, *120*(3), 254-262.
13. Nijkamp, P. G., Habets, L. L., Aartman, I. H., & Zentner, A. (2008). The influence of cephalometrics on orthodontic treatment planning. *The European Journal of Orthodontics*, *30*(6), 630-635.
14. Devereux, L., Moles, D., Cunningham, S. J., & McKnight, M. (2011). How important are lateral cephalometric radiographs in orthodontic treatment planning?. *American Journal of Orthodontics and Dentofacial Orthopedics*, *139*(2), e175-e181.
15. You, W., Hao, A., Li, S., Wang, Y., & Xia, B. (2020). Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health*, *20*(1), 141.
16. Olivetti, E. C., Nicotera, S., Marcolin, F., Vezzetti, E., Sotong, J. P., Zavattero, E., & Ramieri, G. (2019). 3D soft-tissue prediction methodologies for orthognathic surgery—a literature review. *Applied Sciences*, *9*(21), 4550.
17. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.
18. 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*, 106-111.
19. Jeong, S. H., Yun, J. P., Yeom, H. G., Lim, H. J., Lee, J., & Kim, B. C. (2020). Deep learning based discrimination of soft tissue profiles requiring orthognathic surgery by facial photographs. *Scientific reports*, *10*(1), 16235.
20. Sedaghat, B., & Heidari, A. (2026). Artificial intelligence-driven orthodontics: current applications, emerging trends, and key limitations. *Discover Artificial Intelligence*.
21. Yıldırım, A., Cicek, O., Genç, Y. S., & Erdem, R. (2026). Comparative analysis of generative pre-trained transformers for text-and image-based cephalometric prompts using a novel Artificial Intelligence Based Diagnosis and Treatment Planning Index (AIDTI). *BMC Medical Informatics and Decision Making*.
22. Mikulewicz, M., & Chojnacka, K. (2025). Artificial Intelligence in Malocclusion Diagnosis: Capabilities, Challenges, and Clinical Integration. *Applied Sciences*, *15*(24), 13138.
23. Li, Z., & Wang, L. (2025). Multi-task reinforcement learning and explainable AI-Driven platform for personalized planning and clinical decision support in orthodontic-orthognathic treatment. *Scientific Reports*, *15*(1), 24502.
24. Özel, M. B., Kartbak, S. B. A., & Çakmak, M. (2025). Classification Performance of Deep Learning Models for the Assessment of Vertical Dimension on Lateral Cephalometric Radiographs. *Diagnostics*, *15*(17), 2240.
25. Ito, S., Mine, Y., Urabe, S., Yoshimi, Y., Okazaki, S., Sano, M., ... & Tanimoto, K. (2024). Prediction of a cephalometric parameter and skeletal patterns from lateral profile photographs: a retrospective comparative analysis of regression convolutional neural networks. *Journal of Clinical Medicine*, *13*(21), 6346.
26. Surendran, A., Daigavane, P., Shrivastav, S., Kamble, R., Sanchla, A. D., Bharti, L., ... & Shrivastav Sr, S. (2024). The future of orthodontics: deep learning technologies. *Cureus*, *16*(6).
27. Dipalma, G., Inchingolo, A. D., Inchingolo, A. M., Piras, F., Carpentiere, V., Garofoli, G., ... & Inchingolo, F. (2023). Artificial intelligence and its clinical applications in orthodontics: a systematic review. *Diagnostics*, *13*(24), 3677.
28. Kiełczykowski, M., Kamiński, K., Perkowski, K., Zadurska, M., & Czochrowska, E. (2023). Application of artificial intelligence (AI) in a cephalometric analysis: a narrative review. *Diagnostics*, *13*(16), 2640.
29. Li, H., Xu, Y., Lei, Y., Wang, Q., & Gao, X. (2022). Automatic classification for sagittal craniofacial patterns based on different convolutional neural networks. *Diagnostics*, *12*(6), 1359.
30. Rashmi, S., Murthy, P., Ashok, V., & Srinath, S. (2022). Cephalometric skeletal structure classification using convolutional neural networks and heatmap regression. *SN Computer Science*, *3*(5), 336.