Original Research

Title: Performance evaluation of deep learning models for overbite classification on cephalometric radiographs

Running Title: Deep learning for overbite classification

Authors: Sultan Büşra Ay Kartbak[1], Mehmet Birol Özel[2], Muhammed Çakmak[3]

Affiliations: [1]Dr, Kocaeli University, Faculty of Dentistry, Department of Orthodontics, Kocaeli, Türkiye
[2]Dr, Kocaeli University, Faculty of Dentistry, Department of Orthodontics, Kocaeli, Türkiye
[3]Dr, Sinop University, Faculty of Engineering and Architecture, Department of Computer Engineering, Sinop, Türkiye

Received: 01.05.2025

Revised: 05.05.2025

Accepted: 26.06.2025

DOI: 10.26650/eor.20251561000

Authors' ORCID:

Sultan Büşra Ay Kartbak 0000 0003 3862 5658

Mehmet Birol Özel 0000 0002 2984 9468

Muhammed Çakmak 0000 0002 3752 6642

Corresponding author: Mehmet Birol Özel, birolozel@kocaeli.edu.tr

How to cite: Ay Kartbak SB, Özel MB, Çakmak M. Performance evaluation of deep learning models for overbite classification on cephalometric radiographs. Eur Oral Res 2025. Advance online publication.

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

Abstract

Purpose: The objective of this study was to evaluate and compare the effectiveness of different deep learning algorithms in classifying overbite based on lateral cephalometric radiographic images.

Materials and Methods: This study was conducted using lateral cephalometric radiographs of 1062 patients. Overbite values were measured via WebCeph, and the radiographs were categorized into three groups (Overbite 1, Overbite 2, and Overbite 3) based on cephalometric measurements. Six deep learning models (ResNet101, DenseNet201, EfficientNetV2-B0, ConvNetBase, EfficientNet-B0, and a Hybrid Model) were employed to classify the radiographs. Model performance was evaluated using various metrics, including F1-score, accuracy, precision, recall, mean absolute error (MAE), Cohen's Kappa coefficient, and area under the ROC curve (AUC-ROC). Additionally, confusion matrices and Grad-CAM visualizations were generated to further interpret the models' decision-making processes.

Results: All deep learning models employed in this study achieved classification accuracies exceeding 85%. Among them, the EfficientNet B0 and Hybrid models yielded the highest accuracy rates, whereas the ConvNetBase model demonstrated the lowest performance in terms of classification accuracy.

Conclusion: The findings of this study highlight the potential of deep learning models to accurately and reliably classify cephalometric overbite categories without the need for conventional cephalometric analysis.

Keywords: artificial intelligence, deep learning, overbite, cephalometry, orthodontic assessment

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

Introduction

Artificial intelligence (AI) encompasses computer programs designed to perform tasks that traditionally require human cognitive abilities. Initially, AI systems were developed as expert systems, replicating human decision-making processes through conditional logic statements (1).

Artificial neural networks, the foundational units of AI, comprise neurons that typically integrate multiple inputs, which are mathematically processed via nonlinear activation functions. Deep learning (DL), a subset of AI, employs artificial neural networks structured through multiple stacked layers. Convolutional neural networks (CNNs), a specialized type of DL algorithm, feature intricate hidden layer architectures, making them particularly effective for image analysis (2).

In recent years, significant advancements in machine and deep learning have driven their integration into various consumer technologies, including smartphones, cameras, and web search engines. Simultaneously, the digital transformation of healthcare—encompassing the creation of electronic health records and imaging data—has expanded the application of machine and deep learning methodologies in medical and dental practices (1)

Orthodontic datasets hold substantial promise for AI applications due to their standardized, longitudinal, and multimodal structures. Numerous AI-driven solutions have been introduced within the field of orthodontics, with one of the most extensively studied areas being the automated detection of lateral cephalogram landmarks. Traditionally, these landmarks have been used to calculate specific parameters essential for orthodontic treatment planning and assessment (3). Likewise, lateral cephalograms have been employed in classifications of cervical vertebrae maturation stages, decision making for orthognathic surgery and extraction, predictions of upper airway obstructions and degenerative temporomandibular joint diseases (4-10).

The identification of orthodontic treatment need relies on standardized indices that integrate and summarize a broad array of diagnostic data, often emphasizing tooth position

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

analysis (3). Overbite, a crucial component of occlusion, has been a central focus of orthodontic interventions since the field's early days. Overbite correction also serves as an outcome measure for evaluating the quality of orthodontic treatment. Variations in overbite have been linked to skeletal differences among individuals (11).

Although there are studies in the literature that perform automatic skeletal classification with DL algorithms without anatomical landmark marking and cephalometric analysis; to our knowledge, there is no previously published study evaluating overbite with different DL algorithms (12,13). The performance of deep learning algorithms may vary depending on the model and the measured parameter. The aim of this study was to compare the performance of deep learning algorithms in overbite assessment on lateral cephalometric radiographs. The findings of this study are expected to support the development of systems that enhance clinical diagnosis and decision-making, ultimately fostering the creation of more accurate and purpose-specific AI-based diagnostic tools in orthodontics.

Materials and Methods

Ethics statement

This retrospective study was approved by the Ethics Research Committee of the Kocaeli University (Protocol No. KU GOKAEK-2024/14/40)

Data description and preprocessing

Lateral cephalometric radiographs of 1062 patients who applied to Kocaeli University Faculty of Dentistry between 2018-2025 were collected retrospectively from the department archive. Patients with a history of prior orthodontic treatment, craniofacial anomalies, or cleft lip and palate were excluded from the study.

Radiographic evaluation

All the radiographs were taken by the same device (J. Morita MFG. Corp Veraviewepocs 2D Kyoto, Japan) under standardized exposure parameters (80 kV, 10 mA, and 7.4 seconds). A standardized protocol was employed wherein the patients were

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

positioned in the cephalostat with the Frankfort horizontal plane parallel to the floor, the sagittal plane at a right angle to the path of the X-ray, and the teeth in centric occlusion. Cephalometric analysis was performed by a single investigator with 6 years of experience in cephalometric measurement and evaluation, utilizing the WebCeph AI-based orthodontic and orthognathic online platform (AssembleCircle Corp., Gyeonggi-do, Republic of Korea). Magnification correction was performed by referencing a known distance of 10 mm between two fixed points on the cephalostat rod within the radiograph. Following the automatic landmark digitization performed by the program, the landmarks were manually adjusted. To assess intra-operator reliability, measurements were repeated 2 weeks after the initial assessment using records from 50 randomly selected subjects. Intraclass correlation coefficients were calculated using R software (version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria) and were found to range between 0.89 and 0.98, indicating a high level of reproducibility.

Three linear landmarks were identified on each cephalogram and overbite was measured: incisal edges of the maxillary and mandibular central incisors (U1 and L1) and mandibular molar mesial cusp (L6) (Figure 1).

Overbite were obtained from the distance between incisal border of maxillary central incisor to mandibular central incisor and measured on a line perpendicular to occlusomandibular plane (a line bisecting the incisor overbite and passing through the lower molar cusps. (Figure 1).

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

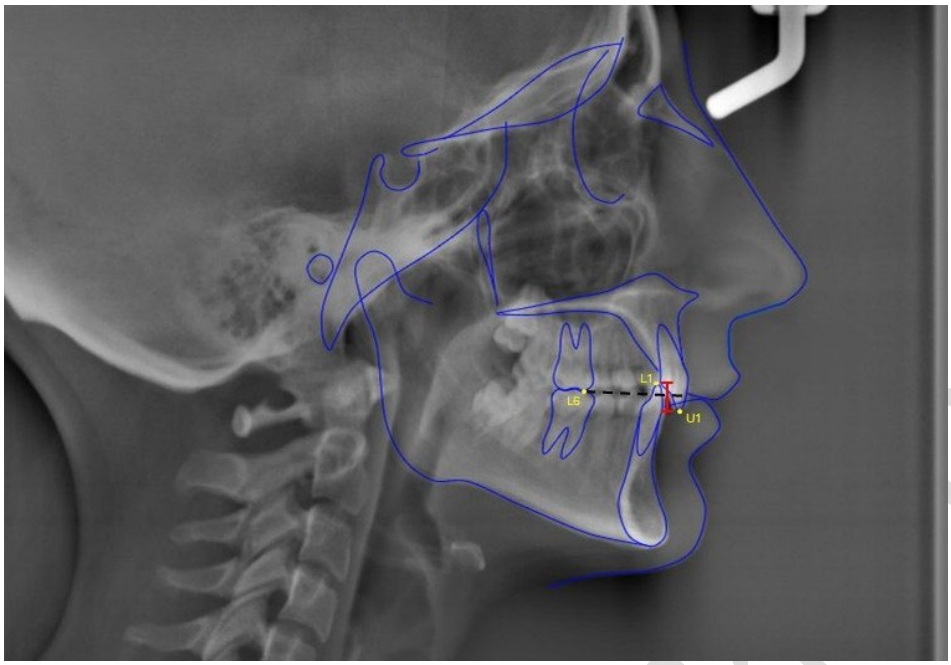


Figure 1. *Cephalometric reference points and overbite measurement used in the study.*

Based on the overbite measurement, the cephalometric radiographs were classified into three subgroups: Overbite 1, Overbite 2, and Overbite 3. To prevent imbalance between the groups, numerical equality was established rather than relying on cephalometric norm values.

The overbite measurements in the Overbite 1 group varied between -10.5 mm and 1.11 mm, encompassing a total of 352 radiographic images; Overbite 2 group varied between 1,12 and 3,11 mm encompassing a total of 354 radiographic images; Overbite 3 group varied between 3,12 and 14,09 mm, encompassing a total of 356 radiographic images.

Image augmentation

To enhance the diversity and robustness of the dataset, several image augmentation techniques were employed. These included horizontal translation, rotation, width and height shifting, as well as zooming operations. Through these augmentation strategies, each subclass was expanded to contain 800 images. After augmentation, the dataset was split

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

into training and testing subsets, where all deep learning models utilized labeled samples. Specifically, 80% of the dataset was allocated for training purposes, while the remaining 20% was reserved for model evaluation.

Model architecture and training details

In this research, six deep learning (DL) architectures were implemented: ResNet101, DenseNet201, EfficientNet B0, EfficientNet V2 B0, ConvNetBase, and a custom Hybrid model. The proposed Hybrid model combines the strengths of EfficientNet B0 and DenseNet201, forming a unified classification framework. Input images of size 224×224×3 were concurrently processed by both networks. EfficientNet B0, recognized for its efficient parameter usage, produced a 1280-dimensional feature vector by extracting high-level representations. In parallel, DenseNet201 generated a 1920-dimensional feature vector by leveraging dense connectivity to capture intricate feature relationships. These vectors were concatenated to create a comprehensive feature space of 3200 dimensions.

To refine feature representation, a Squeeze-and-Excitation (SE) attention mechanism was incorporated, allowing the network to prioritize salient channels and suppress less informative ones. This channel-wise attention facilitated improved learning by reinforcing critical features. The final classification layer was implemented via a multi-layer perceptron (MLP), consisting of progressively reduced layers with 1024, 512, and 256 neurons, culminating in the final prediction output. Regularization and optimization techniques such as Batch Normalization, ReLU activation functions, and Dropout were applied between layers to ensure model stability and prevent overfitting

The training and evaluation procedures were carried out on a GPU-enabled system hosted on Google Cloud, utilizing an NVIDIA Tesla T4 GPU (2.20 GHz) and an Intel Xeon CPU with 16 GB RAM. Python 3 and Keras 2.3.1 were used as the development environment to facilitate model implementation and transfer learning procedures.

Statistical analysis and performance evaluation

Model performance was assessed using several evaluation metrics, including diagnostic accuracy, sensitivity, specificity, area under the ROC curve (AUC), Cohen's

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

Kappa coefficient, and mean absolute error (MAE). In addition, confusion matrices and Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations were generated to provide further insights into the models' interpretability and localization capabilities during training and testing phases.

Results

Cephalometric radiographs were divided into three subgroups as Overbite 1, Overbite 2 and Overbite 3 and the number of samples for each group was increased to 800 by using data augmentation methods. A total of 480 radiographs, 160 from each group, were used in the test dataset of deep learning (DL) algorithms. The classification performance of ResNet101, DenseNet201, EfficientNetV2-B0, ConvNetBase, EfficientNet-B0 and Hybrid algorithms on overbite groups was evaluated. In this context, accuracy, recall, precision and F1 score were measured and the obtained results are presented in Tables 1.

Table 1. *Classification performance of DL algorithms.*

Model	Sınıf	Precision	Recall	F1-Score	Support
ResNet101	Overbite 1	0.9195	0.8562	0.8867	160
	Overbite 2	0.7965	0.8562	0.8253	160
	Overbite 3	0.9057	0.9000	0.9028	160
	Accuracy			**0.8708**	480
DenseNet201	Overbite 1	0.8814	0.9750	0.9258	160
	Overbite 2	0.8741	0.7812	0.8251	160
	Overbite 3	0.9000	0.9000	0.9000	160
	Accuracy			**0.8854**	480
EfficientNet V2 B0	Overbite 1	0.9470	0.8938	0.9196	160
	Overbite 2	0.7802	0.8875	0.8304	160
	Overbite 3	0.9320	0.8562	0.8925	160
	Accuracy			**0.8792**	480
ConvNetBase	Overbite 1	0.9262	0.8625	0.8932	160

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

	Overbite 2	0.7600	0.8313	0.7940	160
	Overbite 3	0.8782	0.8562	0.8671	160
	Accuracy			**0.8500**	480
EfficientNet B0	Overbite 1	0.9091	0.9375	0.9231	160
	Overbite 2	0.8313	0.8625	0.8466	160
	Overbite 3	0.9463	0.8812	0.9126	160
	Accuracy			**0.8938**	480
Hybrid	Overbite 1	0.9091	0.9375	0.9231	160
	Overbite 2	0.8462	0.8250	0.8354	160
	Overbite 3	0.9057	0.9000	0.9028	160
	Accuracy			**0.8875**	480

Confusion matrices were generated for each deep learning model (Figure 2). When the classification performances were compared, the EfficientNet-B0 model correctly classified 429 out of 480 radiographs; the Hybrid model 426, the DenseNet101 model 425, EfficientNet-B0 V2 422, ResNet101 418 and ConvNetBase 408.

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

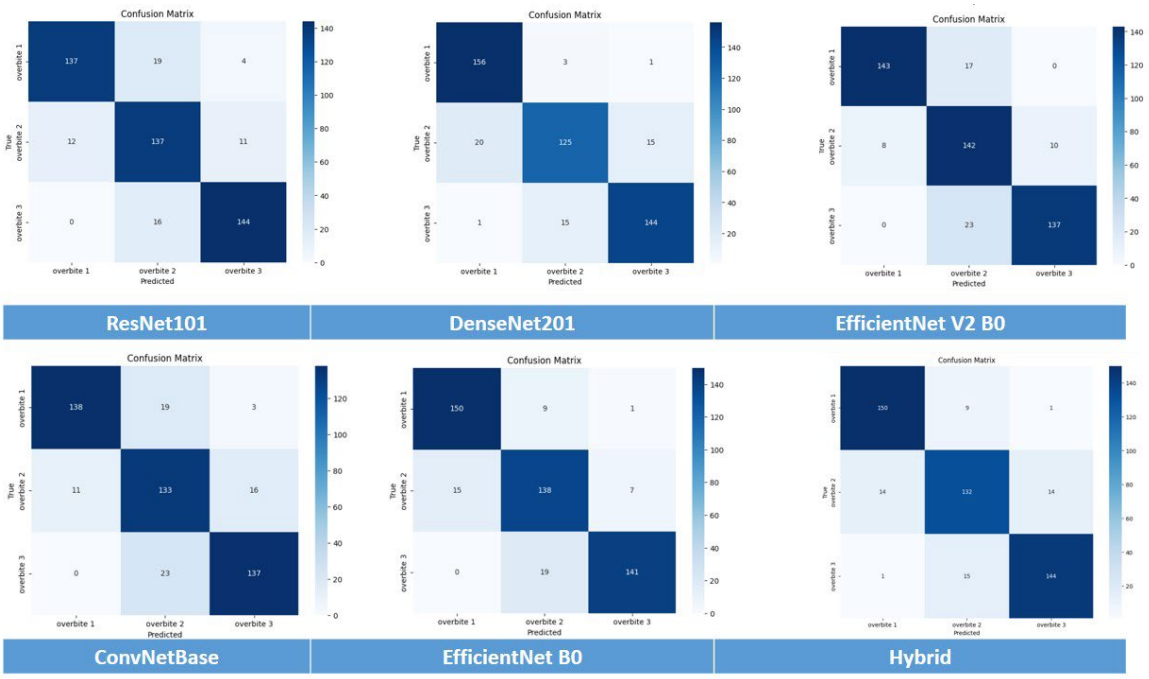


Figure 2. *Confusion matrices of DL algorithms.*

ROC curves were created according to the success of the DL models in classifying the overbite groups and AUC values were determined (Figure 3, Table 2). The highest AUC value was found as 0.9897 for the overbite 1 group in the Hybrid model, while the ConvNetbase model gave the lowest AUC value of 0.9381 for the overbite 2 group.

UNEDITED

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

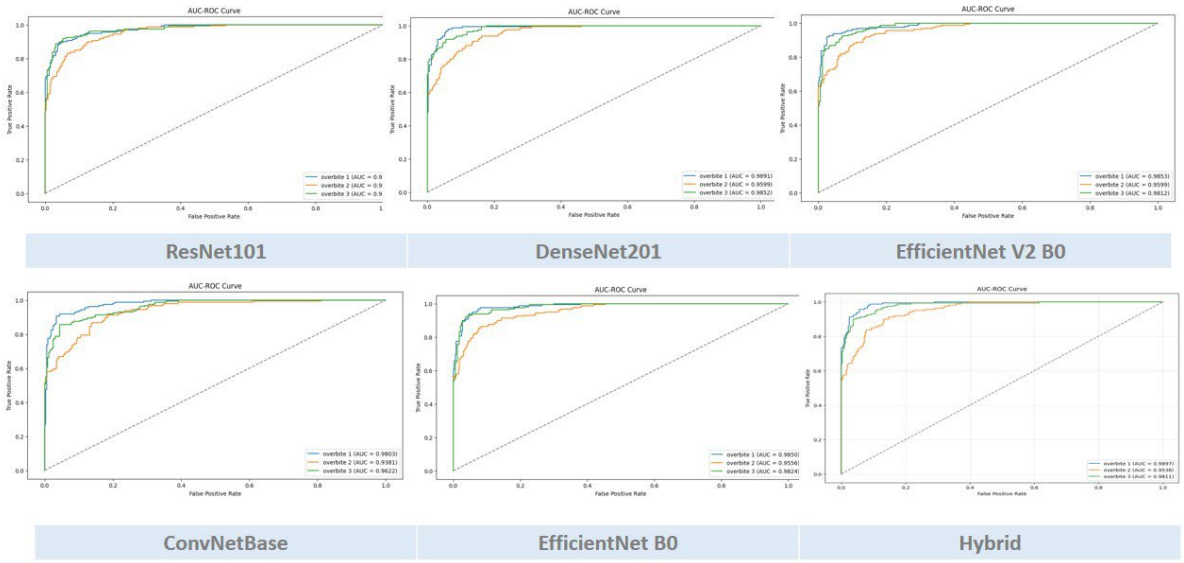


Figure 3. *ROC curves of DL algorithms.*

Table 2. *AUC values of DL algorithms.*

Model	Overbite 1	Overbite 2	Overbite 3
ResNet	0.9765	0.9592	0.9765
DenseNet 201	0.9803	0.9599	0.9852
EfficientNet V2 B0	0.9853	0.9599	0.9812
ConvNetBase	0.9803	0.9381	0.9622
EfficientNet B0	0.9850	0.9556	0.9824
Hybrid	0.9897	0.9538	0.9811

The Table 3 summarizes the accuracy, MAE and Cohen Kappa metrics of deep learning models. EfficientNet B0 (89.38%) has the highest accuracy rate, while ConvNetBase (85.00%) is the model with the lowest accuracy. When evaluated in terms of Mean Absolute Error (MAE), EfficientNet B0 (0.1083) has the lowest error rate, while

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

ConvNetBase (0.1562) is the model with the most errors. Cohen's Kappa metric determined that EfficientNet B0 (0.8406) is the most reliable model in terms of classification consistency, while ConvNetBase (0.7750) is the model with the lowest consistency. Additionally, Figure 4 contains Grad-CAM images of deep learning models. This image provides a visualization of the areas that deep learning models focus on while performing analysis.

Table 3. *Performance metrics of DL algorithm.*

Model	Accuracy	MAE	Cohen's Kappa
ResNet101	0.8708	0.1375	0.8063
DenseNet201	0.8854	0.1187	0.8281
EfficientNetV2 B0	0.8792	0.1208	0.8187
ConvNetBase	0.85	0.1562	0.775
EfficientNet B0	0.8938	0.1083	0.8406
Hybrid	0.8875	0.1167	0.8313

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

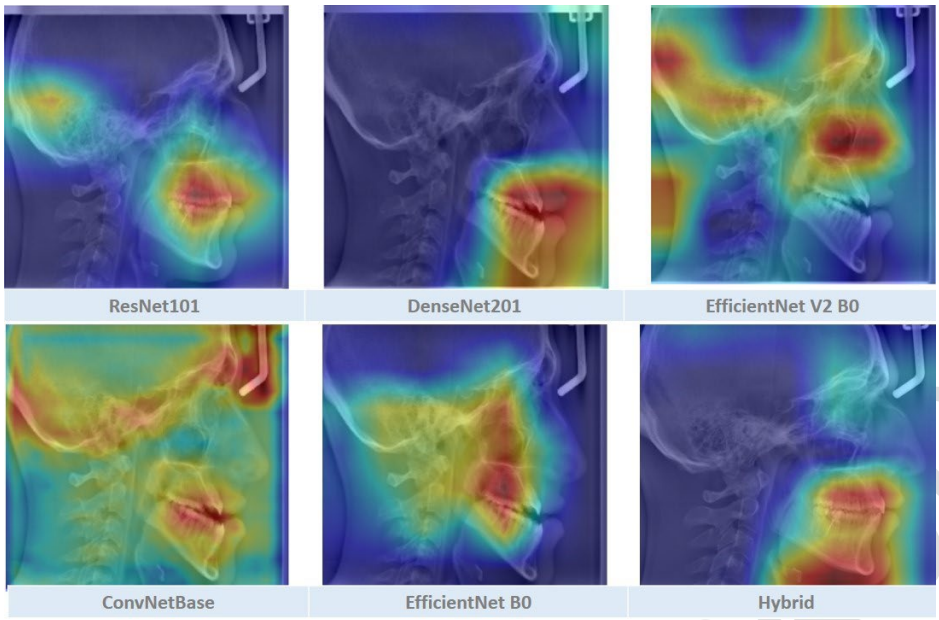


Figure 4. *Grad-CAM visualizations of DL algorithms.*

Discussion

Artificial intelligence is transforming the conventional practices of dentistry by introducing advanced technologies. AI-based systems are predominantly utilized in the development of automated software solutions that enhance diagnostic accuracy and streamline data management in dental care (11).

In recent years, the ability of deep learning to effectively utilize digitized data from large-scale radiological image datasets has led to a growing number of studies focused on the analysis of radiological images using deep learning techniques (14). Convolutional Neural Networks (CNNs), a specialized architecture of deep neural networks for image data, emulate the functionality of the visual cortex by extracting local patterns such as edges and lines, thereby enabling the generation of rich visual features. As a result, CNNs demonstrate superior performance in tasks involving image-based input data (15). The application of deep learning to radiological imaging can generally be categorized into three main tasks: classification, detection, and segmentation (15,16). In our study, we utilized the classification capabilities of deep learning to evaluate its performance in distinguishing between different overbite groups.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

UNEDITED PROOF

Since the development of the Broadbent-Bolton cephalometer, numerous researchers have analyzed cephalograms by identifying key anatomical landmarks, defining reference lines, and measuring angular relationships and lateral cephalometric radiographs has become a standard tool in orthodontic assessment and treatment planning (17,18). This practice also plays a significant role in the frequent use of lateral cephalometric radiographs in artificial intelligence research. In our study, 1,050 cephalometric radiographs were analyzed. Similarly, previous studies employing deep learning techniques have utilized extensive datasets. For instance, in the cervical vertebral maturation (CVM) assessment study conducted by Muhammed Rahimi *et al* (19), 890 lateral cephalometric radiographs were examined, while the automated sagittal skeletal classification research by Nan L *et al.* (12) analyzed 1,613 lateral radiographs.

In this study, six different deep learning algorithms were employed. The EfficientNet B0 algorithm demonstrated the highest classification accuracy at 89.38%, whereas ConvNetBase exhibited the lowest classification accuracy at 85.00%. The other algorithms achieved performance levels that were quite similar to each other. Given that the AUC values exceed 0.9, we conclude that the applied DL models demonstrate excellent diagnostic performance in identifying overbite, indicating their potential utility from a clinical epidemiological perspective. The consistent high performance across all models is promising in terms of highlighting the potential success of deep learning algorithms in overbite measurement.

With the exception of DenseNet201 and ResNet, the classification accuracies of all algorithms were ranked in descending order as Overbite 1, Overbite 3, Overbite 2. The ResNet model demonstrated equal performance in the Overbite 1 and Overbite 3 groups, whereas the DenseNet201 model performed better in the Overbite 3 group compared to Overbite 1. For all models, the lowest performance was observed in the Overbite 2 group. As the values deviated from the norm, the performance of deep learning models increased. Notably, the Overbite 1 group, which included radiographs of patients with reduced overbite, demonstrated the highest level of success. In their study, Yu HJ *et al*. (13) performed automatic skeletal classification using lateral cephalometric radiographs

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

through deep learning techniques. The vertical skeletal pattern was assessed based on the Björk sum and Jarabak ratio, and the classification was carried out using the DenseNet algorithm. The researchers reported a classification accuracy of 96.40% (13). Consistent with the findings of our study, they also noted that the classification accuracy for hyperdivergent and hypodivergent groups was higher compared to that of the normodivergent group (13). One of the key distinctions between our study and that of Yu HJ *et al.* (13) is that our research provides insights into the identification of more effective diagnostic algorithms through the comparison of six different deep learning (DL) models.

Neural networks are often referred to as "black boxes" due to their inherent complexity and feature learning capabilities, which make their decision-making processes difficult to interpret (16). Model explainability plays a crucial role in assessing the interpretability and reliability of the model, identifying its limitations, and uncovering previously unrecognized or hidden patterns within the data (20). To gain a better understanding of the underlying decision mechanisms of a trained CNN, one can examine the receptive field in the input image that leads to the highest activation. This approach provides insight into the role played by specific feature maps in the model's inference (16). In this study, Grad-CAM visualizations were generated for the deep learning models employed. The results indicated that the ConvNetBase model exhibits broader but less precise activation regions compared to other architectures, whereas DenseNet-201 and EfficientNet-B0 demonstrate more focused and less noisy activation areas within the relevant regions.

As presented in Table 3, the EfficientNet B0 model stands out with the highest accuracy (89.38%) and the lowest MAE (0.1083), whereas DenseNet201 demonstrates strong generalization capacity, exhibiting a high Cohen’s Kappa score (82.81%) along with a low MAE (0.1187). The densely connected architecture of DenseNet, which enables feature reuse, combined with EfficientNet’s compound scaling strategy, contributes to a richer and more balanced feature representation. These observations led us to combine the two models within a hybrid architecture. However, the hybrid model did not consistently outperform the best of the individual models. One possible explanation for this outcome is

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

that the models used in the hybrid structure (DenseNet201 and EfficientNet B0) may exhibit similar learning behaviors and may not sufficiently complement each other in terms of feature diversity. In particular, since both models are deep, highly parametric, and optimized architectures, the contribution gained from their combination might have remained limited.

The study presents certain limitations. The utilization of data collected from a single center raises the potential risk of overfitting. Additionally, employing homogeneous datasets may lead to reduced predictive accuracy when applied to diverse data contexts. Furthermore, lateral cephalometric radiograph parameters are prone to variability across different populations. Conducting multicenter studies would allow for larger and more heterogeneous datasets, facilitating the creation of a balanced dataset and thereby enhancing the generalizability of the findings.

Conclusion

The findings of this study demonstrate the potential of deep learning models to reliably estimate cephalometric overbite classifications in the absence of traditional cephalometric assessment. All DL models employed in our study demonstrated classification accuracies exceeding 85%. The EfficientNet B0 and Hybrid models presented the highest accuracy rates, and ConvNetBase presented the lowest accuracy rates. The highest classification accuracy of the DL models was observed in the Overbite 1 group, which included hypodivergent/reduced overbite cases, whereas the lowest accuracy was noted in the Overbite 2 group, characterized by values closer to the norm.

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

Türkçe Öz: *Derin öğrenme modellerinin sefalometrik radyograflar üzerinden overbite sınıflandırma performanslarının değerlendirilmesi. Amaç: Lateral sefalometrik radyografik görüntüler kullanılarak overbite ölçümlerini sınıflandırmada farklı derin öğrenme algoritmalarının etkinliğini değerlendirmek ve karşılaştırmaktır. Gereç ve Yöntem: Çalışmada, 1.062 hastaya ait yanal sefalometrik radyografiler kullanılmıştır. Overbite değerleri WebCeph aracılığıyla ölçülmüş ve radyografiler sefalometrik ölçümlere göre üç gruba ayrılmıştır (Overbite 1, Overbite 2 ve Overbite 3). Görüntüleri sınıflandırmak amacıyla altı farklı derin öğrenme modeli (ResNet101, DenseNet201, EfficientNetV2-B0, ConvNetBase, EfficientNet-B0 ve Hibrit Model) kullanılmıştır. Model performansı; F1 skoru, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), ortalama mutlak hata (MAE), Cohen's Kappa katsayısı ve ROC eğrisi altında kalan alan (AUC-ROC) gibi çeşitli ölçütler kullanılarak değerlendirilmiştir. Ayrıca, modellerin karar verme süreçlerini daha iyi yorumlayabilmek için karmaşıklık matrisleri (confusion matrices) ve Grad-CAM görselleştirmeleri oluşturulmuştur. Bulgular: Bu çalışmada kullanılan tüm derin öğrenme modelleri %85'in üzerinde sınıflandırma doğruluğu elde etmiştir. Modeller arasında EfficientNet B0 ve Hibrit Model en yüksek doğruluk oranlarına ulaşırken, ConvNetBase modeli sınıflandırma doğruluğu açısından en düşük performansı göstermiştir. Sonuç: Bu çalışmanın bulguları, geleneksel sefalometrik analizlere ihtiyaç duyulmaksızın, derin öğrenme modellerinin sefalometrik overbite kategorilerini doğru ve güvenilir bir şekilde sınıflandırma potansiyelini ortaya koymaktadır. Anahtar Kelimeler: yapay zeka, derin öğrenme, overbite, sefalometri, ortodontik değerlendirme*

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

Ethics Committee Approval: This retrospective study was approved by the Ethics Research Committee of the Kocaeli University (Protocol No. KU GOKAEK-2024/14/40).

Informed Consent: Participants provided informed consent.

Peer-review: Externally peer-reviewed.

Author contributions: SBAK, MBÖ participated in designing the study. SBAK participated in generating the data for the study. SBAK participated in gathering the data for the study. SBAK, MBÖ, MÇ participated in the analysis of the data. SBAK wrote the majority of the original draft of the paper. MBÖ, MÇ participated in writing the paper. SBAK has had access to all of the raw data of the study. SBAK, MBÖ has reviewed the pertinent raw data on which the results and conclusions of this study are based. SBAK, MBÖ, MÇ have approved the final version of this paper. SBAK, MBÖ, MÇ guarantees that all individuals who meet the Journal's authorship criteria are included as authors of this paper.

Conflict of Interest: The authors declared that they have no conflict of interest.

Financial Disclosure: The authors declared that they have received no financial support.

UNEDITED PROOF

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

References

1. Mohammad-Rahimi H, Rokhshad R, Bencharit S, Krois J, Schwendicke F. Deep learning: A primer for dentists and dental researchers. J Dent 2023;130:104430.

2. Zhang H, Botler M, Kooman JP. Deep Learning for Image Analysis in Kidney Care [published correction appears in Adv Kidney Dis Health. 2023;30:303. doi: 10.1053/j.akdh.2023.02.001.]. Adv Kidney Dis Health 2023;30:25-32.

3. Nordblom NF, Büttner M, Schwendicke F. Artificial Intelligence in Orthodontics: Critical Review. J Dent Res 2024;103:577-84.

4. Seo H, Hwang J, Jeong T, Shin J. Comparison of Deep Learning Models for Cervical Vertebral Maturation Stage Classification on Lateral Cephalometric Radiographs. J Clin Med 2021;10:3591.

5. Akay G, Akcayol MA, Özdem K, Güngör K. Deep convolutional neural network-the evaluation of cervical vertebrae maturation. Oral Radiol 2023;39:629-38.

6. Choi HI, Jung SK, Baek SH, et al. Artificial Intelligent Model With Neural Network Machine Learning for the Diagnosis of Orthognathic Surgery [published correction appears in J Craniofac Surg. 2020;31:1156. doi: 10.1097/SCS.0000000000006531.]. J Craniofac Surg 2019;30:1986-9.

7. Lee K-S, Ryu J-J, Jang HS, Lee D-Y, Jung S-K. Deep Convolutional Neural Networks Based Analysis of Cephalometric Radiographs for Differential Diagnosis of Orthognathic Surgery Indications. Applied Sciences 2020;10:2124.

8. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop. 2016;149:127-33.

9. Fang X, Xiong X, Lin J, Wu Y, Xiang J, Wang J. Machine-learning-based detection of degenerative temporomandibular joint diseases using lateral cephalograms. Am J Orthod Dentofacial Orthop 2023;163:260-71.e5.

10. Jeong Y, Nang Y, Zhao Z. Automated Evaluation of Upper Airway Obstruction Based on Deep Learning. Biomed Res Int 2023;2023:8231425.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record

11. Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry - A systematic review. J Dent Sci 2021;16:508-22.

12. Nan L, Tang M, Liang B, et al. Automated Sagittal Skeletal Classification of Children Based on Deep Learning. Diagnostics (Basel) 2023;13:1719.

13. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. Automated Skeletal Classification with Lateral Cephalometry Based on Artificial Intelligence. J Dent Res 2020;99:249-56.

14. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.

15. Jang HJ, Cho KO. Applications of deep learning for the analysis of medical data. Arch Pharm Res 2019;42:492-504.

16. Chartrand G, Cheng PM, Vorontsov E, et al. Deep Learning: A Primer for Radiologists. Radiographics 2017;37:2113-31.

17. Taub PJ. Cephalometry. J Craniofac Surg 2007;18:811-7.

18. Devereux L, Moles D, Cunningham SJ, McKnight M. How important are lateral cephalometric radiographs in orthodontic treatment planning?. Am J Orthod Dentofacial Orthop 2011;139:175-81.

19. Mohammad-Rahimi H, Motamadian SR, Nadimi M, et al. Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study. Korean J Orthod 2022;52:112-22.

20. Linse C, Alshazly H, Martinetz T. A walk in the black-box: 3D visualization of large neural networks in virtual reality. Neural Comput Appl 2022;34:21237-52.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record