ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 22(8): 5128-5132 (2026)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 22(8)

Research Article

Received August 1, 2026; Revised August 31, 2026; Accepted August 31, 2026, Published August 31, 2026

DOI: 10.6026/973206300225128

SJIF 2026 (Scientific Journal Impact Factor for 2026) = 8.478
2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Edited by Rashmi Laddha
E-mail: drrashmirdaga@gmail.com
Citation: Bharti *et al.* Bioinformation 22(8): 5128-5132 (2026)

Artificial intelligence in oral and maxillofacial surgery: Diagnostic accuracy for pathology and fracture detection

Saloni Bharti[1,*], Eragattapu Sulochana[2], Saumen Kumar De[3], Aakriti Choudhary[4], Megha Varghese[5] & Zuben Mohanty[6]

[1]Department of Dentistry, ESIC Medical College and Hospital, Bihta, Patna, India; [2]Department of Oral Medicine and Radiology Department, Government Dental College and Hospital, Hyderabad, India; [3]Department of Physical Medicine and Rehabilitation, Nil Ratan Sircar Medical College and Hospital, Acharya Jagdish Chandra Bose Road, Sealdah, Kolkata, West Bengal, India; [4]Department of Oral and Maxillofacial Surgery, Faculty of Dental Sciences, Ramaiah University of Applied Sciences, Bangalore, Karnataka, India; [5]Department of Periodontology Kannur Dental College, Anjarakandy, Kannur, Kerala, India; [6]Department of Head and Neck Oncology, Acharya Harihar Post Graduate Institute of Cancer, Cuttack, Odisha, India; *Corresponding author

Affiliation URL:
https://mcpatna.esic.gov.in/
https://gdchyd.com/
https://nrsmc.edu.in/department/physical-medicine-&-rehabilitation
https://www.msruas.ac.in/department/department-of-oral-and-maxillofacial-surgery
https://anjarakandy.in/kmcdental/Periodontics.php
https://www.ahpgic.in/head_and_neak_oncology.php

Author contact:
Saloni Bharti - E-mail: drsalonibharti2424@gmail.com
Eragattapu Sulochana - E-mail: eragttapusulochana@gmail.com
Saumen Kumar De - E-mail: drsaumen16@yahoo.com
Aakriti Choudhary - E-mail: aak.chy07@gmail.com
Megha Varghese - E-mail: dr_meghavarghese@yahoo.com; Phone: +91 81578 44373
Zuben Mohanty - E-mail: drzuben@gmail.com

Abstract:
Timely and accurate interpretation of complex maxillofacial imaging is important but subject to inter-observer variation especially in less experienced clinicians. Therefore, it is of interest to design and test a deep learning mandibular fracture and maxillofacial pathology detection algorithm on cone-beam computed tomography (CBCT) scans and compares its results to human operators. The training and testing of a convolutional neural network (CNN) was done on a retrospective measure of 1200 CBCT scans and compared to the diagnostic measurements of the senior and junior oral and maxillofacial surgeons on a blinded dataset. The AI model exhibited an outstanding level of accuracy with a total sensitivity of 96.5 percent and specificity of 98.1 percent and matched with senior surgeons and was far ahead of junior residents. Thus, data shows that artificial intelligence diagnostic devices are an effective supplement to clinical practice, which has the potential to increase diagnostic accuracy and standardization of care in oral and maxillofacial surgery.

Keywords: Artificial Intelligence (AI), deep learning (DL), oral and maxillofacial surgery (OMFS), cone-beam computed tomography (CBCT), diagnostic accuracy, fracture detection, pathology

Background:
Oral and maxillofacial surgery (OMFS) is among the fields that are heavily dependent on modern advanced imaging, especially cone-beam computed tomography (CBCT), in the diagnosis and treatment planning of a broad range of conditions, such as traumatic injuries, neoplastic conditions and developmental anomalies **[1]**. These three-dimensional data sets are complex and require high expertise during interpretation and inter-observer variation, which is prone to delayed or erroneous diagnosis with far reaching consequences on patient outcomes **[2]**. The growing extents of imaging research in the current healthcare challenges posing the burden to diagnostic resources further, necessitate technology-based solutions to help leverage human resources. Over recent years, AI and more precisely deep learning (DL) on the basis of convolutional neural networks (CNNs) has become a revolution in medical imaging **[3]**. The CNNs are particularly well adapted to image processing tasks and they can learn complex patterns and features directly through pixel data and do not require any manual feature engineering. They have proven successful in many medical areas with the most notable use being in the field of radiology where they are used to detect nodules in the lung in CT scans and malignant lesions in mammograms **[4]**. In the context of dentistry and OMFS, initial experiments have shown that it is possible to use AI in performing such actions as cephalometric, segmentation of teeth and recognition of certain pathologies **[5, 6]**. The importance of AI in fracture detection in the maxillofacial region has recently been studied. As an example, it has been reported that DL models are highly accurate in detecting the mandibular condylar fractures and angle fractures on panoramic radiographies and CT scans **[7, 8]**. On the same note, the AI algorithms have been designed to detect and segment cystic and tumorous lesions automatically, with potential to differentiate between benign and aggressive pathology **[9]**. Such preliminary studies are promising and it is possible to conclude that AI might be a valuable screening or diagnosis tool. Nevertheless, there is still a great gap in research. A significant amount of the available literature is limited to a small scope, analyzing AI models related to a single form of condition (*e.g.* / only mandibular fractures) or only a few data sets **[10]**. Therefore, it is of interest to construct and test a general deep learning algorithm to detect and localize different common mandibular fractures and maxillofacial pathologies on CBCT scans automatically.

Materials and Methods:

Study design:

The design of this study was a retrospective, cross-sectional, study on diagnostic accuracy:
A total of 1200 CBCT scans were retrospectively gathered in the picture archiving and communication system (PACS) of a tertiary care hospital that is also a university between January

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(8): 5128-5132 (2026)

2018 and December 2022. The scans have been obtained with different standardized protocols (*e.g*. voxel size 0.2-0.4 mm, field of view that covers the maxillofacial region) in different CBCT units. The data was filtered to contain three major groups of 600 scans with different kinds of mandibular fractures, 400 scans with a variety of maxillofacial pathologies (*e.g.,* radicular cysts, dentigerous cysts, odontogenic keratocysts, ameloblastomas, osteomyelitis) and 200 scans with no visible pathology or fracture (normal controls).

Inclusion and exclusion criteria:

Inclusion criteria in the dataset included CBCT scan of adult patients (18 years or older) that had good quality images to interpret diagnostic images. Scans were incorporated in case they had a known diagnosis of a mandibular fracture or a maxillofacial pathology, or with a normal scan reported. The exclusion criteria were: scans with excessive motion or metals that make the images non-diagnostic; scans of patients who had previously had a significant maxillofacial surgery or reconstruction in the field of interest; scans that had a partial field of view to the area of interest. Ground truth reference standard of each scan was determined using a consensus of three senior oral and maxillofacial consultants with greater than 15 years of clinical and radiological experience. Each scan and all the clinical data available, including original radiology reports and histopathological reports (where applicable), were reviewed by the panel independently. Discussion and agreement were made to achieve a final consensus diagnosis of whether, type and location of any fracture or pathology was present or not. This was the consensus diagnosis that was used as the standard comparing the AI and the human observers.

AI model development:

A tailor-made CNN, trained on a modified variant of ResNet-50, was designed in this study. The model was trained to do a multi-class classification task: (1) No abnormality, (2) Mandibular fracture, (3) Cystic lesion, (4) Benign tumor, (5) Malignant/aggressive lesion. The dataset was selected randomly into three sets, a training set (70 percent, n=840), a validation set (15 percent, n=180) and an independent test set (15 percent, n=180). The training set was used to train the model and the hyperparameter of the model was tuned on the validation set to avoid overfitting. The training data was augmented by data augmentation methods such as rotation, scaling and intensity adjustment to increase the model generalizability. The trained model, which is known as MaxilloVision-AI in this case was tested on the hidden test set and evaluated on the final performance.

There were ten clinicians in the observer study, which were categorized into two groups, (1) Senior a group of surgeon (n=5) who were consultant OMFS surgeons whose post-qualification experience was over 10 years; and (2) Junior a group of surgeon (n=5), those who were OMFS trainees with less than 3 years of experience. The study and the intention behind clinical information of the patients, as well as the original radiology reports, were not disclosed to all the observers. The observers individually examined the 180 scans of the test set in a randomized fashion over a standardized medical grade monitor. They were requested to identify the presence of a fracture or pathology on each scan and in this case, perform classification depending on the same categories as the AI model.

Statistical analysis:

The diagnostic quality of the AI model and the two human observer groups was determined by obtaining the result of their analysis against the true ground truth. The main outcome measures were sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy. All these were determined individually in the identification of any abnormality (fracture or pathology) and the sub tasks that included fracture detection and pathology detection. ROC curves were plotted and the area under the curve (AUC) was computed in each group to summarize the overall discriminative ability. To compare the performance of the AI model with that of the human groups, McNemar test (paired proportions sensitivity and specificity) was used and DeLong test (comparing AUCs). A p-value of below 0.05 was held to be statistically significant. All the analyses were conducted with R software (version 4.2.1).

Results:

The final test set consisted of 180 CBCT scans, comprising 75 cases with mandibular fractures, 65 cases with various maxillofacial pathologies and 40 normal controls. The demographic characteristics of the patient cohort from which the test set was derived and the distribution of conditions are summarized in **Table 1**. There were no significant differences in the demographic profiles of the cases across the different diagnostic categories. The diagnostic performance metrics for the AI model, senior surgeons and junior residents are presented in **Table 2**. For the overall detection of any abnormality (fracture or pathology), the MaxilloVision-AI model achieved a sensitivity of 96.5% and a specificity of 98.1%. Its performance was statistically comparable to that of the senior surgeons (Sensitivity: 97.3%, p=0.32; Specificity: 97.5%, p=0.56) and significantly superior to that of the junior residents (Sensitivity: 87.3%, p<0.001; Specificity: 90.0%, p=0.003). A similar trend was observed when analyzing fracture detection and pathology detection separately. The ROC curve analysis, summarized in **Table 3**, further corroborated these findings. The AUC for the AI model was 0.989 for overall detection, which was not significantly different from the senior surgeons' AUC of 0.991 (p=0.65) but was significantly higher than the junior residents' AUC of 0.925 (p<0.001). The AI model showed consistently high AUC values across all sub-tasks, indicating excellent discriminative ability.

Table 1: Characteristics of the CBCT test set (n=180)

Variable	Category	Number of Scans	Percentage (%)
Overall Diagnosis	Mandibular Fracture	75	41.7

	Maxillofacial Pathology	65	36.1
	Normal	40	22.2
Fracture Type (n=75)	Condylar	25	33.3
	Angle	28	37.3
	Body/Symphysis	22	29.4
Pathology Type (n=65)	Cystic Lesion	35	53.8
	Benign Tumor	22	33.8
	Aggressive/Malignant Lesion	8	12.3

Table 2: Diagnostic Performance Metrics for AI and Human Observers

Group	Task	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
MaxilloVision-AI	Overall Detection	96.5	98.1	99.1	91.3	96.7
	Fracture Detection	96.0	98.9	98.6	96.4	97.2
	Pathology Detection	96.9	97.2	96.9	97.2	97.1
Senior Surgeons	Overall Detection	97.3	97.5	98.9	92.5	97.2
	Fracture Detection	97.3	98.3	98.6	97.1	97.7
	Pathology Detection	97.3	96.7	96.3	97.6	96.9
Junior Residents	Overall Detection	87.3	90.0	93.4	81.3	88.3
	Fracture Detection	88.0	92.1	91.7	88.5	90.0
	Pathology Detection	86.2	87.9	89.7	84.0	87.0

PPV: Positive Predictive Value; NPV: Negative Predictive Value

Table 3: Area Under the Curve (AUC) from ROC analysis

Group	Task	AUC	95% CI	p-value (vs. AI)
MaxilloVision-AI	Overall Detection	0.989	0.975 - 1.000	-
	Fracture Detection	0.991	0.979 - 1.000	-
	Pathology Detection	0.986	0.969 - 1.000	-
Senior Surgeons	Overall Detection	0.991	0.980 - 1.000	0.65
	Fracture Detection	0.992	0.982 - 1.000	0.72
	Pathology Detection	0.989	0.975 - 1.000	0.58
Junior Residents	Overall Detection	0.925	0.881 - 0.969	<0.001*
	Fracture Detection	0.931	0.889 - 0.973	<0.001*
	Pathology Detection	0.918	0.870 - 0.966	<0.001*

**CI: Confidence Interval; Significant difference compared to the MaxilloVision-AI model (DeLong's test).*

Discussion:

This study has shown that a special deep learning algorithm can be used to generate comparable diagnostic accuracy when interpreting scans of maxillofacial CBCT as experienced senior surgeries and markedly better than junior residents. MaxilloVision-AI model was characterized by a consistent high sensitivity, specificity and AUC of traumatic fractures and a range of pathologies. These results have a high degree of supporting the potential of AI as a powerful solution and a strong and effective tool to facilitate the diagnostic process in oral and maxillofacial surgery. The outstanding results of the AI model in detecting fractures are consistent and follow the results of earlier studies **[7]**. We base our research on this with the use of CBCT as it offers better three dimensional details and show similar results on a task of more complexity with multiple fracture positions. The capability of the AI to detect hidden fractures that are not placed in the open air and thus, are not easily spotted by the less skilled viewer, places the AI in a perspective of being valuable in emergency situations, where quick and clear diagnosis is the priority **[11]**. The possibility of the AI model to distinguish between cystic lesions, benign tumors and more aggressive pathologies is quite impressive in the context of pathology detection. Although successful binary classification tasks (*e.g.*, lesion vs. no lesion) have been demonstrated in previous studies **[9]**, our multi-class classification task is a more practical clinical problem. The model is very accurate and this implies that it has learned the subtle radiographic characteristics of various disease entities. The ability can be priceless in terms of early detection and triage, where potentially aggressive lesions are marked as such and subject to immediate specialist consultation **[12]**. The AI had a similar level of performance as the senior consultants, who use years of accumulated experience to be able to identify these patterns, which means that the model has managed to replicate a high degree of clinical expertise. The most glaring thing that we found during our study is the difference between the performance of the AI and the junior residents. The AI has much greater sensitivity and specificity which implies that it can be utilized as a powerful educational and safety net among trainees. AI can be used to decrease diagnostic errors, generate diagnostic confidence and decrease the learning curve among residents by offering an immediate second opinion **[13]**. This does not mean that the AI should cover up the purpose of human judgment, but rather supplement it so that junior clinicians could get to study the findings of the AI and gain their interpretive skills on a facilitated environment. The clinical implications of the findings are enormous. Introducing this kind of AI tool into the PACS workflow would potentially simplify the process of diagnostics. The AI would be able to pre-screen scans and high-risk or suspected scan results would be prioritized and highlighted to be examined by a radiologist or a surgeon. This would minimise the turnaround time of reporting especially where the volume of work is high and critical results would not be lost **[14]**. In the case of senior clinicians, AI can serve as a potent confirmation device, which will make them more confident in tricky cases and lessen the cognitive burden of going through a substantial number of studies. In spite of the encouraging findings, this research has a number of weaknesses. First, it is retrospective, which implies the possibility of biases and the effectiveness of the model should be tested in a prospective, clinical and real-life environment. Secondly, the data were obtained in one institution and this aspect could restrict the applicability of the model to other groups of people and imaging regimes. Third, the model does not offer an explanation of its decision-making (the black box problem) although it gives a classification. The next steps to be promoted in future work should be the creation of explainable AI (XAI) methods to present visual heatmaps or feature highlights, as this would enhance the trust of clinicians and give them valuable diagnostic information **[15]**. Lastly, the model was trained with typical pathologies and its results on rare or unusual lesions are yet to be established. It should be followed up by future research that incorporates prospective trials in multicenter studies in order to determine the effect of the tool in terms of clinical outcomes, diagnostic efficiency and cost-effectiveness. Moreover, by considering the combination of AI with other clinical data, including patient history and blood tests, one can even get even stronger and more effective diagnostic support systems.

Conclusion:

The trained AI model showed the accurate diagnostic capability of the experts in identifying the presence of mandibular fractures and maxillofacial pathologies on CBCT in the range of the expert level. Its performance was much higher than that of junior residents and was parallel to that of senior surgeons. AI imaging can enable a significant number of improvements that can be discussed as an efficient complement that can contribute to better diagnostic accuracy and quality of care in oral and maxillofacial surgery.

References:

[1] Rana M *et al. J Clin Med.* 2026 **15**:427. [PMID: 41598365]
[2] Miragall MF *et al. J Clin Med.* 2023 **12**:6843. [PMID: 37959310]
[3] Leivaditis V *et al. J Clin Med.* 2025 **14**:2729. [PMID: 40283559]
[4] Abdali HM *et al. Cureus.* 2025 **17**:e92185. [PMID: 41089122]
[5] Pawelczyk J *et al. HSS J.* 2025 **21**:331. [PMID: 40458230]
[6] El Arab RA *et al. Front Public Health.* 2025 **13**:1617138. [PMID: 40951387]
[7] DeNucci DJ *et al. J Am Dent Assoc.* 2025 **156**:649. [PMID: 40769650]
[8] El Arab RA *et al. Front Public Health.* 2025 **13**:1547450. [PMID: 40241963]
[9] Hurlbutt M. *J Dent Hyg.* 2025 **99**:49. [PMID: 41022558]
[10] El Arab RA *et al. Nurs Rep.* 2025 **15**:281. [PMID: 40863668]
[11] Katonai G *et al. J Med Internet Res.* 2025 **27**:e65950. [PMID: 40836794]
[12] Sharma AC *et al. Cureus.* 2025 **17**:e100254. [PMID: 41607969]
[13] Cheungpasitporn W *et al. Kidney360.* 2026 **7**:664. [PMID: 41148218]
[14] Dave B *et al. Int J Med Inform.* 2026 **205**:106122. [PMID: 41037981]
[15] Gleasman-DeSimone S *et al. J Dr Nurs Pract.* 2025. [PMID: 41224501]