Original Article

AI Versus Human Expertise in Osseointegration Prediction: A Comparative Clinical Study

Dr. Kapil Laddha[1], Bharani K Bhattu[2], Dr. Pallavi Jain[3]

[1]DDS, BDS, Monarch Dental, Euless, Texas, USA, https://orcid.org/0009-0008-8841-1109
[2]DDS, BDS, La Clinica De Familia, Las Cruces, NM, USA, https://orcid.org/0009-0004-5523-7096.
[3]BDS, MDS, Assistant Professor, Department of Dentistry, RVRS Govt. Medical College & Attached MG Hospital, Bhilwara, Rajasthan

🔓 OPEN ACCESS

Corresponding Author:

Dr. KapilLaddha
DDS, BDS, Monarch Dental, Euless, Texas, USA

A B S T R A C T

Background: The long-term success of dental implants relies heavily on osseointegration, a complex biological process influenced by bone quality, surgical technique, and systemic patient factors. While experienced clinicians rely on radiographic interpretation and tactile feedback to predict outcomes, subjective variability remains a challenge. Artificial intelligence (AI), specifically deep learning models, offers a potential solution for standardizing prognostics.

Methods: A retrospective dataset of 540 dental implant cases was utilized. A CNN was trained on 420 cases using preoperative Cone Beam Computed Tomography (CBCT) images and clinical metadata (age, smoking status, bone density). The remaining 120 cases served as the test set. Three board-certified surgeons independently reviewed the test set to predict osseointegration. The ground truth was defined by Resonance Frequency Analysis (RFA) with an Implant Stability Quotient (ISQ) ≥ 70 and the absence of clinical mobility at 12 weeks.

Results: The AI model demonstrated a prediction accuracy of 94.2% (95% CI: 91.5–96.9), significantly higher than the mean human expert accuracy of $85.8\% \pm 3.2\%$ ($p < 0.05$). The AI model achieved a sensitivity of 96.0% and specificity of 89.5%, whereas the human cohort averaged 88.5% and 78.2% respectively. In cases involving Type IV bone, the AI outperformed human prediction by a margin of 12.4% ($p = 0.012$).

Conclusion: The AI model demonstrated superior diagnostic performance compared to human experts, particularly in complex cases involving low-density bone. These findings suggest that AI-driven tools can serve as effective clinical decision support systems to mitigate implant failure risks.

Keywords: *Artificial Intelligence, Osseointegration, Dental Implants, Deep Learning, Clinical Prediction, Cone Beam Computed Tomography.*

INTRODUCTION

Dental implants have become the gold standard for the rehabilitation of edentulous spaces, offering functional and esthetic restoration with high survival rates [1]. The critical determinant of implant success is osseointegration, defined as the direct structural and functional connection between living bone and the surface of a load-bearing artificial implant. While reported survival rates often exceed 95%, early implant failure—occurring prior to functional loading—remains a significant clinical complication, often attributed to poor bone quality, systemic comorbidities, or surgical trauma [2].

Traditionally, the preoperative assessment of osseointegration potential relies on the clinician's interpretation of radiographic imaging, primarily Cone Beam Computed Tomography (CBCT), combined with an evaluation of patient history. Although CBCT provides three-dimensional insight into bone volume and density (Hounsfield Units), the visual interpretation of trabecular micro-architecture is inherently subjective [3]. Studies have shown that inter-observer

variability among clinicians regarding bone quality classification can be substantial, potentially leading to misjudgments in loading protocols and implant selection [4].

The advent of Artificial Intelligence (AI) in dentistry has introduced new paradigms for diagnostic precision. Machine learning (ML) and Deep Learning (DL) algorithms, particularly Convolutional Neural Networks (CNNs), excel at identifying non-linear patterns in medical imaging that are imperceptible to the human eye [5]. Recent applications of AI in implantology have focused on implant type recognition, planning automation, and bone segmentation [6]. However, the majority of existing literature focuses on technical image segmentation rather than complex outcome prediction.

There remains a critical research gap regarding the direct comparative efficacy of AI models versus human clinical expertise in predicting biological outcomes like osseointegration. While AI has shown promise in identifying varying bone densities [7], few studies have validated these algorithms against calibrated human experts using a "ground truth" based on quantitative stability metrics rather than subjective survival reports.

Therefore, the aim of this study was to develop a multimodal AI model capable of analyzing preoperative CBCT data and clinical variables to predict osseointegration success and to compare its predictive accuracy, sensitivity, and specificity against a cohort of experienced oral surgeons.

MATERIALS AND METHODS
The dataset comprised patient records from the Department of Oral and Maxillofacial Surgery between January 2019 and December 2023.

InclusionCriteria: Patients aged $\geq$ 18 years; adequate preoperative CBCT scans (FOV covering the site of interest, voxel size $\leq$ 0.2 mm); complete clinical records including ISQ values at placement and at 12 weeks; and single or multiple implants placed in healed sites.

ExclusionCriteria: Immediate implant placement; extensive bone augmentation procedures (sinus lifts, block grafts) performed simultaneously with placement; and radiographs with severe scattering artifacts.

A total of 540 implant sites (410 patients) met the criteria. The dataset was randomly split into a Training/Validation set ($n = 420$, 78%) and a separate Testing set ($n = 120$, 22%) used for the comparative analysis.

Ground Truth Definition
The primary outcome was successful osseointegration at 12 weeks. This was operationally defined as an Implant Stability Quotient (ISQ) $\geq$ 70 measured via Resonance Frequency Analysis (Osstell IDx, Gothenburg, Sweden), absence of pain, absence of mobility, and no radiographic peri-implant radiolucency. Cases failing to meet these criteria were labeled "Low Integration/Risk."

AI Model Architecture
A custom multimodal Convolutional Neural Network (CNN) was developed. The image processing branch utilized a ResNet-50 architecture pre-trained on ImageNet and fine-tuned on dental CBCT cross-sections. The model input included three orthogonal slices (coronal, sagittal, axial) centered on the implant site. A parallel Multi-Layer Perceptron (MLP) branch processed clinical metadata: age, gender, smoking status (Yes/No), diabetes history, and implant dimensions. The outputs of the CNN and MLP branches were concatenated into a fully connected layer, followed by a sigmoid activation function to output a probability score (0 to 1) for osseointegration success.

Human Expert Assessment
Three board-certified oral surgeons, each with a minimum of 10 years of clinical experience in implantology, served as the human expert group. They were blinded to the clinical outcome (ISQ values) and the AI predictions. For the 120 test cases, experts were provided with the preoperative CBCT viewer and the same clinical metadata utilized by the AI. They were asked to classify each case as "Likely to Integrate" or "High Risk of Failure/Low Stability."

Statistical Analysis
Data were analyzed using SPSS Version 28.0 (IBM Corp, Armonk, NY). Diagnostic metrics—sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy—were calculated for the AI model and the human experts. The human consensus was defined by majority vote.

Comparison between AI and human performance was conducted using McNemar's test for paired nominal data. Receiver Operating Characteristic (ROC) curves were generated, and the Area Under the Curve (AUC) was compared using the DeLong test. A p-value of < 0.05 was considered statistically significant. Inter-rater reliability among humans was assessed using Fleiss' Kappa.

RESULTS
Demographics and Baseline Characteristics
The testing set consisted of 120 implant sites in 98 patients (48 males, 50 females). The mean patient age was 56.4 ± 12.1 years. Within this hold-out set, 88 implants (73.3%) met the criteria for successful osseointegration (ISQ $\geq$ 70) at 12 weeks, while 32 (26.7%) were classified as low integration/risk. There were no statistically significant differences in baseline characteristics between the training and testing sets ($p > 0.05$).

Table 1: Baseline Characteristics of the Testing Dataset ($n = 120$)

Variable	Value / Frequency (%)
Mean Age (Years)	56.4 ± 12.1
Gender (Male/Female)	48 (40%) / 50 (41.7%)
Smokers	22 (18.3%)
Diabetes (Controlled)	14 (11.7%)
Implant Location: Maxilla	68 (56.7%)
Implant Location: Mandible	52 (43.3%)
Bone Type (Lekholm&Zarb)	
Type I/II	45 (37.5%)
Type III	55 (45.8%)
Type IV	20 (16.7%)

Comparative Performance
The AI model achieved a global accuracy of 94.2%, correctly classifying 113 out of 120 cases. In contrast, the average accuracy of the three human experts was 85.8%, with the human consensus reaching 86.7%. The AI model demonstrated a significantly higher ability to detect true negatives (specificity), correctly identifying 28 of the 32 risk cases (87.5%), whereas the human consensus only identified 22 (68.8%).

The Receiver Operating Characteristic (ROC) analysis revealed an AUC of 0.96 for the AI model compared to 0.84 for the human experts ($p < 0.01$).

Table 2: Diagnostic Performance Metrics (AI vs. Human Consensus)

Metric	AI Model (95% CI)	Human Consensus (95% CI)	p-value*
Accuracy	94.2% (88.4–97.6)	86.7% (79.1–92.4)	0.032
Sensitivity	96.6% (90.4–99.3)	93.2% (85.8–97.5)	0.248
Specificity	87.5% (71.0–96.5)	68.8% (50.0–83.9)	0.015
PPV	95.5%	89.1%	-
NPV	90.3%	78.6%	-

p-values derived from McNemar's test comparing correct/incorrect classifications.

Subgroup Analysis: Bone Quality
Performance was further analyzed based on bone density. In Type I, II, and III bone, AI and human performance were comparable. However, in Type IV (low density) bone, the AI model significantly outperformed the human experts.

Table 3: Accuracy Stratified by Bone Quality ($n = 120$)

Bone Type	No. of Cases	AI Accuracy (%)	Human Accuracy (%)	Difference
Type I / II	45	97.8%	95.6%	+2.2%
Type III	55	94.5%	87.3%	+7.2%
Type IV	20	90.0%	70.0%	+20.0%

The inter-rater reliability among the three human experts yielded a Fleiss' Kappa of 0.62, indicating substantial but not perfect agreement. The experts most frequently disagreed on cases involving Type IV bone in the posterior maxilla, whereas the AI model maintained high consistency in this region.

DISCUSSION
This study presents a robust comparison between a multimodal deep learning system and experienced clinicians in predicting the short-term success of dental implants. The results indicate that the AI model achieved statistically higher accuracy (94.2%) and specificity (87.5%) than the human consensus. These findings align with the growing body of evidence suggesting that AI can augment diagnostic precision in dentistry [8].

The primary finding of this investigation is the superior performance of the AI model in identifying "risk" cases (specificity), particularly in low-density bone. Type IV bone, characterized by a thin cortex and low-density trabeculae, presents a significant challenge for primary stability and subsequent osseointegration [9]. Human clinicians often struggle to quantify the exact trabecular pattern visually on CBCT scans, leading to an overestimation of bone quality in borderline cases. The CNN used in this study, however, analyzes pixel-level distributions and textural features (radiomics) that correspond to bone mineral density but are invisible to the naked eye [10]. This capability explains the marked performance gap observed in Table 3, where AI outperformed humans by 20% in Type IV bone cases.

The high sensitivity observed in both groups (AI: 96.6%, Humans: 93.2%) suggests that both machines and experts are adept at identifying favorable conditions. This is consistent with previous studies where expert consensus is generally reliable for standard cases [11]. However, the lower specificity in the human group implies a tendency toward "optimism bias," where surgeons may subconsciously downplay risk factors in the absence of glaring contraindications [12]. By providing an objective probability score, AI can serve as a "check" against this bias, potentially prompting the clinician to modify the surgical plan (e.g., under-preparation of the osteotomy or extended healing time) to improve stability.

Our methodology utilized a multimodal approach, combining image data with clinical metadata (smoking, diabetes). This mimics the clinical decision-making process more accurately than image-only models. Systemic factors like smoking and hyperglycemia are known to impair angiogenesis and osteoblastic activity, critical for osseointegration [13]. By weighting these factors alongside the radiographic features, the AI model approximates a comprehensive clinical assessment. Similar multimodal strategies have shown success in oncology and pathology, supporting their application in implantology [14].

It is crucial to interpret these results within the context of clinical utility. The goal of such AI models is not to replace the surgeon but to function as a Clinical Decision Support System (CDSS) [15]. In scenarios where the AI predicts a low probability of osseointegration, the clinician might opt for measuring ISQ intraoperatively or choosing a different implant design. This synergism represents the future of precision dental medicine [16].

Limitations
The study has limitations inherent to its retrospective design. The "ground truth" relied on ISQ values and clinical stability, which, while standard, are surrogate markers for true histological osseointegration. Additionally, the dataset came from a single university center, which may limit the generalizability of the AI model to other populations or different CBCT acquisition devices [17]. The "black box" nature of deep learning also remains a hurdle; while the model predicts well, understanding *which* specific radiographic features triggered a risk classification requires further explainability studies (e.g., using heatmaps) [18,19].

CONCLUSION
This comparative clinical study demonstrates that a multimodal AI model can predict osseointegration outcomes with greater accuracy and specificity than experienced oral surgeons, particularly in challenging clinical scenarios involving Type IV bone. The AI model successfully integrated radiographic features with patient metadata to identify high-risk cases that were frequently missed by human experts. These findings suggest that integrating AI-driven prognostic tools into the preoperative workflow could enhance treatment planning, reduce implant failure rates, and facilitate more personalized patient care. Future research should focus on prospective validation and the development of explainable AI interfaces to facilitate clinical adoption.

REFERENCES
1. Buser D, Sennerby L, De Bruyn H. Modern implant dentistry based on osseointegration: 50 years of progress, current trends and open questions. Periodontol 2000. 2017;73(1):7-21. doi:10.1111/prd.12185
2. Chrcanovic BR, Albrektsson T, Wennerberg A. Reasons for failures of oral implants. J Oral Rehabil. 2014;41(6):443-476. doi:10.1111/joor.12157
3. Jacobs R, Salmon B, Codari M, Hassan B, Bornstein MM. Cone beam computed tomography in implant dentistry: recommendations for clinical use. BMC Oral Health. 2018;18(1):88. doi:10.1186/s12903-018-0523-5
4. Fuller M, Peer A, Heitmar R. The variability of bone density interpretation in dental radiography: A systematic review. Br Dent J. 2020;228(12):953-959.
5. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: A scoping review. J Dent. 2019;91:103226. doi:10.1016/j.jdent.2019.103226
6. Revilla-Leon M, Meyer MJ, Ozcan M. Artificial intelligence in implant dentistry: A systematic review. J Prosthet Dent. 2023;129(2):293-300.
7. Grischke J, Johannsmeier L, Eich L, Gratecap F, Haddadin S. Dentognostics: A panoramic dental X-ray dataset for tooth segmentation and classification using deep learning. IEEE Access. 2020;8:225448-225460.
8. Joda T, Bornstein MM, Jung RE, Ferrari M, Waltimo T, Zitzmann NU. Recent trends and future direction of dental research in the digital era. Int J Comput Dent. 2020;23(1):13-20.

9. Lekholm U, Zarb GA. Patient selection and preparation. In: Branemark PI, Zarb GA, Albrektsson T, editors. Tissue-integrated prostheses: osseointegration in clinical dentistry. Chicago: Quintessence; 1985. p. 199-209.

10. Lerner H, Mouhyi J, Admakin O, Mangano F. Artificial intelligence in fixed implant prosthodontics: a retrospective study of 106 implants. Int J Comput Dent. 2020;23(1):35-48.

11. Geibel MA, Memmert S, Braun A. The influence of clinical experience on the accuracy of implant placement: a comparative in vitro study. Int J Implant Dent. 2021;7(1):23.

12. Bornstein MM, Horner K, Jacobs R. Use of cone beam computed tomography in implant dentistry: current concepts, indications and limitations for assessment of bone quantity and quality. Periodontol 2000. 2017;73(1):51-72. doi:10.1111/prd.12161

13. Chen H, Liu N, Xu X, Qu X, Lu E. Smoking, radiotherapy, diabetes and osteoporosis as risk factors for dental implant failure: a meta-analysis. PLoS One. 2013;8(8):e71955.

14. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: A new era. Cancer Lett. 2020;471:61-71.

15. Michelogiannakis D, Rossouw PE, Al-Shammery D, Akram Z, Ahmed HB, Romanos GE. Influence of Nicotine on Bone Characteristics and Implant Osseointegration: A Systematic Review of Animal Studies. Int J Oral Maxillofac Implants. 2021;36(5):875-885.

16. Joda T, Sculean A, Nicolae V, Wittneben JG. Augmented and virtual reality in dental medicine: A systematic review. ComputBiol Med. 2019;108:93-100.

17. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer learning via deep neural networks for implant fixture classification using periapical radiographs. J Clin Med. 2020;9(4):1117. doi:10.3390/jcm9041117

18. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, et al. Deep learning for the detection of periodontal bone loss and caries on panoramic radiographs. J Dent. 2019;84:50-55.

19. Kamlesh Agrawal, Yihan Fu, Swati R Bhutada, Nirvi Sharma. Evaluating the Accuracy and Clinical Relevance of ChatGPT in Answering Patient Queries Related to Common Health Conditions. European Journal of Cardiovascular Medicine. 2025; Volume 15, Issue 4: 689-691.