Use of Convolutional Neural Networks for Detection of Pathologies in Dental X-Ray Images in Clinical Decision Support Systems

Sviatoslav Dziubenko[1,*], Andriy Kyrylyuk[2], Volodymyr Krasnov[3], Valentyn Avakov[4] and Oksana Atamanchuk[5]

[1]*Department of Information and Communication Technologies named after O.O. Zelensky, Faculty of Radio Electronics, Computer Systems and Infocommunications, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine*

[2]*Department of Dentistry, PO, Ivano-Frankivsk National Medical University, Ivano-Frankivsk, Ukraine*

[3]*Interregional Academy of Personnel Management, Kyiv, Ukraine*

[4]*Department of Pediatric Dentistry, Faculty of Dentistry, Ivano-Frankivsk National Medical University, Ivano-Frankivsk, Ukraine*

[5]*Department of Histology, Cytology and Embryology, Ivano-Frankivsk National Medical University, Ivano-Frankivsk, Ukraine*

Abstract: *Relevance:* The relevance of the study is determined by the need for automated, scalable solutions for processing large volumes of dental radiological images, which provide precise segmentation, detection, and classification of pathologies in the integrated Clinical Decision Support System (CDSS) modules.

Aim: The aim of the study is to develop, optimize, and verify a HITL-CDSS framework for dental radiology with multi-level integration of Convolutional Neural Network (CNN) models, ensuring architectural consistency, metric validity, and expert adaptability.

Methods: Research methods: critical architectural and functional analysis of CNN models, metric and indicator modelling of efficiency, synthesis and Unified Modelling Language-based (UML)modelling of the CDSS framework, UML optimization with Human-in-the-loop (HITL) integration, metric and indicator verification of HITL-CDSS.

Results: Architectural and functional, metric and indicator, as well as UML modelling of CNN architectures was carried out for the purpose of integration into the dental radiology CDSS. The resultant HITL-optimized framework based on DenseNet/EfficientNet, HRNet, YOLOv8 provided AUC = 0.96–0.98, F1@t = 0.91–0.94, DSC = 0.89–0.92, mAP = 0.72–0.77 at ECE = 0.02–0.04. Integration of HITL mechanisms increased Explainable Artificial Intelligence (XAI) interpretability, resistance to domain shifting, and clinical validity, indicating the appropriateness of multi-modular construction of CDSS with the inclusion of expert feedback.

Conclusion: The academic novelty of the study is the development of a HITL-CDSS framework with multi-level CNN integration, which provides metrically verified interpretability, domain-stable generalizability, and clinical relevance in dental radiology tasks.

Keywords: Medical data, neural networks, neural network training, dentistry, x-rays, image processing, object detection.

INTRODUCTION

Modern dental radiology is undergoing transformation under the influence of deep learning (DL) methods, in particular CNNs, which demonstrate high potential in the tasks of segmentation, detection, and classification of oral pathologies. At the same time, the integration of CNN architectures into CDSS requires architectural and functional analysis, metric verification, and adaptation to the real clinical environment. The concept of HITL is gaining particular relevance, which enables increasing the clinical and diagnostic relevance of CDSS by involving the expert knowledge of dental radiologists in the training cycle.

The aim of the study is to develop, optimize, and verify a HITL-oriented CDSS framework for dental radiology based on multi-level integration of CNN models, ensuring architectural consistency, metric and indicator validity, as well as expert adaptability in a clinical context.

Research objectives:

- perform an architectural and functional analysis of CNN models regarding their integration suitability in CDSS;
- perform normalized metric and indicator modelling with visualization in Heatmap format;
- synthesize a UML CDSS model with modular integration of relevant CNN architectures;
- optimize the architecture through HITL mechanisms: active learning, expert feedback, and adaptive reconfiguration;

*Address correspondence to this author at the National Aerospace University "Kharkiv Aviation Institute", Vadyma Manka Str., 17, Kharkiv, Ukraine, 61070; E-mail: standardct2@gmail.com

- perform verification of HITL-CDSS by discriminative, segmentation, and detection metrics.

Current and relevant academic approaches to the use of CNNs in dental radiology for the detection of pathologies on X-ray images in clinical decision support systems are considered. The emphasis is on architectural and algorithmic innovations, segmentation, classification solutions and integration models that shape the modern technical and methodological landscape of the industry.

Starting from a generalized historical and methodological context, Sum [1] presented a chronological overview of the integration of AI into dentistry, where CNN architectures have provided high-precision radiological detection, cephalometric landmarking and oncoscreening. The current stage includes predictive analytics, NLP charting and AI-assisted robotic surgery with an emphasis on ethical and regulatory compliance.

Developing this broad perspective, Shekarappa *et al*.[2] summarized the integration of AI in dentistry, emphasizing the use of ML and DL algorithms, including CNN, for high-precision analysis of radiographs, computed tomography (CT) scans, and intraoral scans. The technologies increase diagnostic precision, predictive analytics, and personalization of therapeutic protocols, reducing errors and optimizing clinical workflows.

Deepening the thematic focus, Akram [3] reviewed the transformations of periodontology under the influence of AI, in particular CNN architectures (U-Net, Mask R-CNN, VGG-16) for segmentation and classification of radiographs and salivary biomarkers. The methodology increases diagnostic sensitivity, prognostic specifics and the level of personalization of the periodontal treatment planning process, taking into account ethical and legal regulations.

In the field of diagnostic and visualization solutions, Hong *et al*. [4] developed a CNN-assisted visualization and analysis platform for processing panoramic radiographs, which integrates intelligent diagnostics, automated extraction of diagnostic descriptors, and statistical correlation analysis of dental nosologies. The architecture increases diagnostic productivity, informativeness of electronic medical records, as well as visual and analytical reports to support public dental health strategies.

Deepening the DL-diagnostics aspect, Nour *et al*.[5] developed an integrated CNN architecture with YOLO8+RT-DETR detector ensemble, optimized by Non-Maximum Suppression method, which achieved mAP50 = 74% and mAP50–90 = 58% in multi-class detection of dental pathologies. The model outperformed the competitors, increasing the accuracy by 30% and reducing the processing time by 18%, ensuring high diagnostic reliability.

Regarding highly specialized segmentation tasks, Firincioglulari *et al*. [6] found that CNN-centric DL-segmentation architectures provide highly accurate identification of pulp (Dice=0.84; IoU=0.758) and pulp stones (Dice=0.759; IoU=0.686) on panoramic radiographs. The models are characterized by increased classification and segmentation precision and recall, confirming their adjuvant clinical and diagnostic relevance.

In the explainable diagnostics, Asghar *et al*.[7] found that CariesXplainer, a CNN-centric AI–XAI architecture combining MobileNetV3 with Grad-CAM in a transfer learning loop, achieves 99.50% caries classification accuracy, significantly outperforming the SOTA CNN architecture. The methodology provides high-level diagnostic interpretability and spatial localization precision.

As regards multi-architecture integrations, Parkhi *et al*. [8] demonstrated that a CNN–ResNet–ViT classification model for five dental nosologies achieves 87.6% validation accuracy using confidence-thresholding and an expert validation module. The system improves diagnostic reproducibility, clinical interoperability, and accessibility in resource-constrained environments.

In the context of a critical methodological assessment, Mahizha *et al*.[9] found that systematic DCNN architectures of the segmentation-classification-detection type for the identification of interproximal caries on bitewing radiographs provide diagnostic preference of YOLOv8. Only 40% of the analysed papers had a low risk of bias of the reference standard according to QUADAS-2.

Concluding the analysis, Singh *et al*. [10] systematized the application of CNN-oriented DL-architectures in caries and prosthodontic classification, covering tooth shade selection, restoration design, and classification of morphofacial changes in patients with removable dentures. The ability of the models to process heterogeneous unstructured data while maintaining classification constraints was noted.

The analysis of the reviewed publications confirms that CNNarchitectures provide high diagnostic precision, sensitivity, and specifics in multi-class

detection of dental pathologies, significantly increasing the interpretability and speed of radiological examinations. Further automation of diagnostics in dental radiology requires in-depth optimization of CNN architectures by reducing computational complexity, increasing robustness to visual artifacts and inter-device variability, and integrating explainable AI and attention mechanisms to ensure traceable interpretation, as well as clinical and metric validation of diagnostic decisions.

METHODS AND MATERIALS

Research design

The study was carried out in successive iterations – Figure **1**.

Methods

The study employed the following methods:

1. *Critical architectural and functional analysis*: a systematic assessment of CNN architectures (Table **1**) was carried out for their discriminatory ability, computational efficiency, and ability for modular integration into the CDSS framework of dental radiology.

2. *Metric and indicator modelling*: a comparative normalized evaluation of CNN models by metrics (Table **2**) was implemented with the construction of Heatmap and Performance Score aggregation for multi-criteria interpretation of efficiency.

3. UML-based synthesis modelling of the CDSS framework: a UML structure of the synthesized CDSS framework was built with the integration of the most relevant CNN components, determined based on the previous modelling phases, for the implementation of a modular diagnostic system.

4. UML optimization (hybrid HITL-CDSS): the architecture was improved by including HITL elements: expert training mechanisms, active sampling, UI-interface for review and adaptive optimization learning cycle.

5. Metric-indicator verification of HITL-CDSS: quantitative verification of the hybrid framework (HITL-CDSS) was carried out with the construction of a Heatmap profile, which showed an improvement in discriminative (AUC = 0.96–0.98) and segmentation (DSC = 0.89–0.92) characteristics, confirming the effectiveness of HITL integration.

Sample

A systemic selection of verified CNNs tested in dental radiology for automated classification, detection and segmentation of pathologies is presented below (Table **1**). These architectures, developed taking into account the specifics of x-ray images of the dento-maxillofacial system, cover algorithmic principles, structural features and clinical scenarios for their integration into CDSS to improve diagnostic accuracy, speed, and interpretability of results.

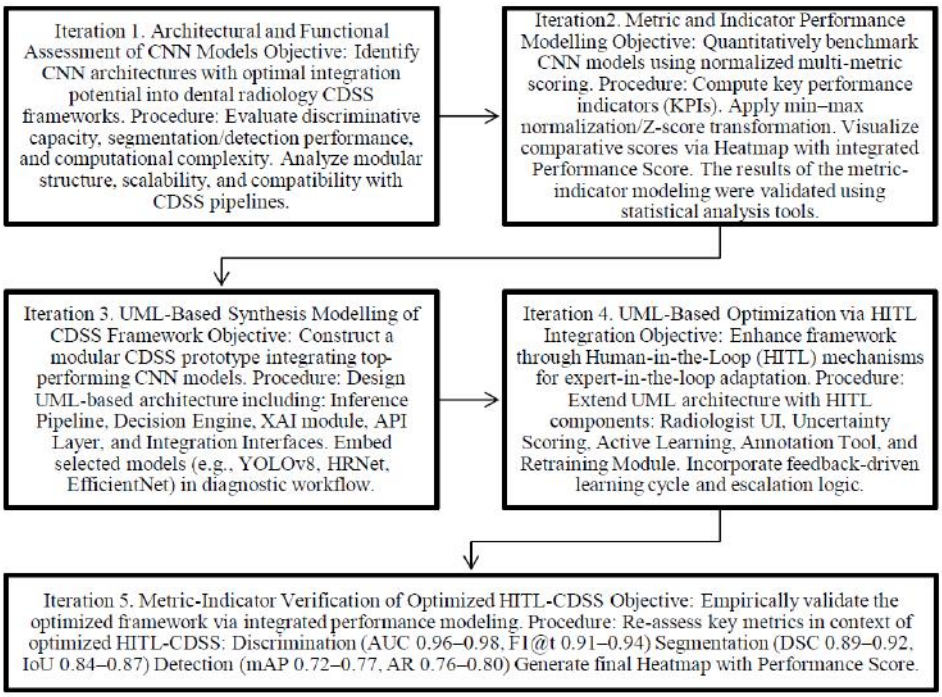

Figure 1: Multi-iterative research design.

Source: developed by the authors.

Table 1: Case of Verified Convolutional Neural Network Architectures for Automated Detection of Dental Pathologies in X-Ray Images

CNN model name / Task type (evaluation metrics)	Brief description	Architecture / functional features	Verified examples of application in dentistry	Academic research
U-Net / Segmentation (DSC (Dice Similarity Coefficient), IoU, Pixel Accuracy, Sensitivity, Specificity)	Symmetric encoder–decoder segmentation network with skip-connections	Architecture with contraction and expansion paths; use of symmetric convolutional blocks; multi-scale context / High accuracy of pixel segmentation; resistance to data limitations; adaptability to medical images	Segmentation of caries, cysts, periapical lesions (University of Freiburg (2015))	[11, 12]
Mask R-CNN /Segmentation + Detection (mAP, IoU, DSC, Precision, Recall, F1-score)	Extension of Faster R-CNN with mask segmentation	RPN (Region Proposal Network) + ROIAlign architecture for accurate localization; parallel branches for classification, detection and segmentation / High localization accuracy; integrated multitasking	Detection and segmentation of caries, tumours, bone defects (Facebook AI Research (2017))	[13, 14]
YOLOv5 / YOLOv8 / Detection (mAP@0.5, mAP@0.5:0.95, Precision, Recall, Inference Time)	High-performance single-stage object detectors	CSPDarknet backbone, PANet neck, adaptive anchor-free mode (YOLOv8); integration of auto-learning anchors / Real-time operation; multi-class detection; optimization of inference time	Automated detection of caries, pulp stones, periodontal lesions (Ultralytics (2020 / 2023))	[15, 16]
ResNet-50 / ResNet-101 / Classification (Accuracy, AUC-ROC, Precision, Recall, F1-score)	Deep CNN with residual blocks (skip-connections)	Bottleneck blocks with identity mapping; elimination of accuracy degradation with increasing depth / Stable learning even with >100 layers; high AUC and F1 metrics	Classification of pathologies on panoramic and bitewing images (Microsoft Research (2015))	[17, 18]
DenseNet-121 / DenseNet-169 / Classification (Accuracy, AUC-ROC, Precision, Recall, F1-score)	Densely connected layer network	Each layer receives input from all previous ones; improved gradient propagation / Fewer parameters with high accuracy; efficient on small datasets	Detection of early stages of caries, periodontitis (Cornell University, Tsinghua University, Facebook AI Research (2017))	[19, 20]
MobileNetV3 / Classification / Detection (Accuracy, mAP, Inference Time, Model Size)	Lightweight CNN for mobile systems	Inverted residual blocks + squeeze-and-excitation; optimization for low-FLOPs computing / Low latency; energy efficiency; deployment in portable-CDSS	Mobile diagnostics and telemedicine in dentistry (Google AI (2019))	[21, 22]
Inception-v3 / Inception-ResNet-v2 / Classification (Accuracy, AUC-ROC, Precision, Recall, F1-score)	Multi-scale Feature Processing Networks	Factorized convolutions; parallel convolutional filters of different sizes; Inception-ResNet – added residual connections / Efficiency with different-sized objects; improved classification of complex pathologies	Classification and stratification of pathologies (Google Research (2015 / 2016))	[23, 24]
EfficientNet (B0–B7) / Classification / Detection (Accuracy, AUC-ROC, mAP, Inference Time, Params)	Optimally Scalable CNN	Compound scaling (depth, width, resolution); Mobile Inverted Bottleneck Convolution (MBConv) / High accuracy/resource-intensive; easily adaptable	Automated classification of dental radiographs (Google Brain (2019))	[25, 26]
HRNet / Segmentation/Classification (DSC, IoU, Accuracy, Sensitivity, Specificity)	High-Resolution Network for Segmentation	Preserving high resolution at all stages; parallel convolutional branches with cross-scale fusion / High segmentation detail; excellent reproduction of small structures	Localization of micropathologies, segmentation of dental structures (Microsoft Research Asia (2019))	[27, 28]

Note: *CNN architectures in dental radiology were trained and tested by using open repositories of medical radiological images were used, accumulating anonymized orthopantomograms (OPG), bitewing and periapical radiographs, and cone-beam computed tomography (CBCT) with verified diagnostic annotations. The training base covered a wide range of pathologies of the dentofacial system, including carious lesions of various localization, periapical granulomas and cysts, periodontal defects with alveolar bone resorption, pulp stones, pulp calcifications, as well as benign and malignant neoplasms and developmental anomalies. The involvement of open databases (Dental X-ray Images Dataset (Kaggle), Panoramic Dental X-rays Dataset (Mendeley Data), Medical Radiography Open Database (AIcrowd/Grand Challenge), Dental Caries and Lesion Dataset (UCI Repository)) ensured the representativeness of pathological scenarios, standardization of mark-up, and a unified basis for comparative evaluation of CNN models in the context of their integration into clinical decision support systems.

Source: developed by the authors.

Table 2: Instrumental Case of CNN Model Evaluation Metrics*

Metric	Short definition	Mathematical formula
Classification case (ResNet, DenseNet, Inception, EfficientNet, MobileNetV3, HRNet-cls)		
Primary discrimination$(AUC-ROC)$	Discrimination ability by ROC curve	$AUC=\int_0^1 TPRdFPR$ where TPR – completeness of pathology detection:: $TPR=Se=\frac{TP}{TP+FN}$ where TP – rue positive detections; FN – false negative detections; FPR – inverse correctness of the norm cut-off: $FPR=1-Sp=\frac{FP}{FP+TN}$ where FP – false positive detections; TN – true negative detections
F1-score (operating point) $(F1@t)$	Balance between selectivity of positive decisions (P) and completeness of pathology detection (Se)	$F1@t=2\times\frac{P\times Se}{P+Se}$ where P – selectivity of positive decisions: $P=\frac{TP}{TP+FP}$ t – поріг, що визначається за Youden's: $J=Se+Sp-1$ where Sp – correctness of the norm cut-off: $Sp=\frac{TN}{TN+FP}$
Expected Calibration Error (ECE)	Average confidence and precision mismatch (calibration)	$ECE=\sum_{m=1}^{M}\frac{\lvert B_m\rvert}{n}\lvert acc(B_m)-conf(B_m)\rvert$ where B_m – m^{th} bin; n – the total number; $acc(B_m)$ – the proportion of correct answers in the bin; $conf(B_m)$ – average confidence
Segmentation case (U-Net, HRNet-seg, Mask R-CNN-mask)		
Dice Similarity Coefficient (DSC)	Primary planar similarity	$DSC=\frac{2\lvert A\cap B\rvert}{\lvert A\rvert+\lvert B\rvert}$ where A – predicate mask; B – standard
Intersection over Union (IoU)	Mask overlap fraction	$IoU=\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$
Hausdorff Distance $(HD95)$	Marginal error. $HD95$approximation is used – 95$^{\text{th}}$Hausdorff percentile	$HD=\max\left\{\sup_{a\in A}\inf_{b\in B}d(a,b),\sup_{b\in B}\inf_{a\in A}d(a,b)\right\}$ where $d(\cdot,\cdot)$ – Euclidean distance
Detection case (YOLOv5/YOLOv8, Mask R-CNN-det, EfficientNet/MobileNetV3-det)		
Medium Average Precision $(mAP@[0{,}5:0{,}95])$	Primary localization accuracy	$mAP@[0{,}5:0{,}95]=\frac{1}{10}\sum_{\tau=0{,}5}^{0{,}95}AP@IoU(\tau)$ where AP – Average Precision: area under the PR curve for a class at a given IoU threshold: $AP=\int_0^1 P(R)dR$ where $P(R)$ – precision as a function of completeness; IoU threshold τ fixed mAP – average AP by class/threshold: $mAP=\frac{1}{K}\sum_{k=1}^{K}AP_k$ where K – classes; $\tau\in\{0.5,0.55,\ldots,0.95\}$
Average Recall (AR)	Complementary localization completeness –average completeness by IoU grid/detection limit	$AR=\mathbb{E}_{\tau,N_{det}}[R(\tau,N_{det})]$ where R – recall; τ – IoU threshold; N_{det} – detection limit

(Table 2). Continued

Metric	Short definition	Mathematical formula
Inference Efficiency(t_{inf}), (FPS)	Operational performance (latency)	$t_{inf} = \frac{1}{n}\sum_{i=1}^{n} t_i$ where t_i – processing time of one sample; $FPS = \frac{1}{t_{frame}}$ where t_{frame} – frame time
General case (computational efficiency)		
Parametric Complexity $(Params)$, $(FLOPs)$	$Params$ – number of model weights; $FLOPs$ – number of floating point operations per inference	–
Memory/Size	$Model\ size$ (MB), $VRAM/RAM$in inference	–

Note: * The Performance Score normalization method was used For a generalized possibility of comparative analysis of CNN models. It was performed according to the following algorithm:

a. Each metric was reduced to a unitary scale [0;1] by min–max normalization/z-score transformation, which ensured metric unification and correct intermodel comparability.

b. For each class of tasks (classification, segmentation, detection), weight coefficients were formed based on clinical relevance. General integral:

$$Performance\ Score = \sum_{i=1}^{k} w_i \times M_i$$

where w_i – metric weight; M_i – normalized metric value; k – number of metrics.

To ensure transparency, the weighting scheme was predefined: discriminative metrics (AUC-ROC, F1@t, mAP) received higher weights due to their direct impact on diagnostic validity; segmentation metrics (DSC, IoU, HD95) were weighted moderately to reflect their role in morphometric accuracy; calibration metrics (ECE) and operational metrics (AR, t_inf) were assigned auxiliary weights to capture reliability and deployability. The coefficients were fixed across all models and validated through sensitivity analysis, which demonstrated the stability of ranking outcomes under ±15% perturbations of w_i.

c. The resulting $Performance\ Score$ was interpreted as the generalized performance of the CNN model. d. The stability of the integral indicator was checked through bootstrap resampling at the patient level (BCa intervals), as well as sensitivity analysis for changes in the weights w_i to assess robustness.

Source: developed by the authors.

To address class imbalance–a typical limitation of dental radiology datasets –stratified sampling, weighted loss functions, and balanced mini-batch construction were applied during CNN training. These procedures mitigated disproportionate class representation and reduced bias in F1 and mAP metrics, ensuring stable discriminatory behaviour across minority pathology classes.

Instruments

The choice of metrics (Table **2**) was driven by the need for unified intermodel validation of CNN architectures in dental radiology: discriminant (AUC, $mAP@[0{,}5:0{,}95]$,DSC/IoU) reflect the ability to classify, detect, and segment; operational ($F1_t$, AR) characterize functionality in clinically relevant scenarios; calibration (ECE) ensure the reliability of probabilistic predictions; and deployment-oriented (t_{inf}, $Params$, $FLOPs$,$Size$) capture computational efficiency. Statistical robustness and reproducibility of evaluations were ensured by using DeLong, Wilson CI, and bootstrap methodology with patient-level resampling.

The BCa-bootstrap procedure was implemented with 10 000 resamples, which ensured stable estimation of bias-correction and acceleration parameters across all performance metrics (AUC-ROC, F1@t, DSC, IoU, mAP, AR). Resampling was performed strictly at the patient level, not at the image level, to prevent intra-patient autocorrelation from inflating variance. Each bootstrap iteration drew whole-patient image sets with replacement, preserving the original class distribution and maintaining stratification by pathology type, thereby ensuring clinically valid data balance. Confidence intervals were computed using the bias-corrected and accelerated method, with the acceleration coefficient derived from jackknife influence values on patient-level observations. This procedure provided robust, distribution-free estimates of variability and supported the statistical validity of intermodel comparisons.

Python was used for metric and indicator CNN modelling employing the NumPy, SciPy, scikit-learn libraries (statistical processing, metrics), PyTorch/ TensorFlow (CNN training), Matplotlib/Seaborn (visualization), statsmodels (hypothesis testing, DI). The settings included GPU acceleration (CUDA/cuDNN), stratified cross-validation, bootstrap resampling (BCa), and random seed parameter control. An optimized CDSS framework based on relevant CNN models was developed by using UML-based modelling, which provided the formalization of architectural and functional dependencies, the specification of

interoperable modules, and the structural decomposition of the components of the CDSS.

The study used a comprehensive statistical validation procedure implemented in the Python environment using the scikit-learn, statsmodels, and SciPy libraries. Statistical analysis tools, including hypothesis testing, confidence interval construction, bootstrap resampling, and analysis of variance, were implemented using the delong_roc_test, proportion_confint, bootstrap, anova_lm, pairwise_tukeyhsd, and posthoc_dunn functions from the scikit-posthocs library. This approach ensured the reliability, reproducibility, and statistical validity of the results obtained in the context of metric-indicator comparison of CNN models.

The study used a comprehensive statistical validation procedure implemented in Python using scikit-learn, statsmodels, and SciPy libraries. Statistical analysis tools – including hypothesis testing, confidence interval construction, bootstrap resampling, and analysis of variance – were implemented using the delong_roc_test, proportion_confint, bootstrap, anova_lm, pairwise_tukeyhsd, and posthoc_dunn functions from the scikit-posthocs library. This approach ensured the reliability, reproducibility, and statistical validity of the obtained results in the context of metric and indicator comparison of CNN models.

RESULTS

An architectural and functional analysis of verified CNN models tested in dental radiology was performed in the first iteration of the study (Table **3**). The implementation prospects were assessed in terms of resource efficiency, scalability, resistance to domain shift, as well as clinical and diagnostic relevance to identify optimal application scenarios in automated diagnostics and clinical decision support.

The architectural and functional analysis of CNN models (Table **3**) showed the relevance of U-Net for voxel segmentation, Mask R-CNN for multiclass instance segmentation, YOLOv5/8 for high-speed triage, EfficientNet as an optimized backbone of CDSS, and HRNet for morphometric-precision segmentation. At the same time, none of the architectures provided autonomous integration into CDSS due to performance limitations, parametric complexity, and domain stability.

Table 3: Critical Architectural and Functional Analysis of CNN Models for Integration Capability in the CDSS Framework of Dental Radiology

Model	Advantages of application	Disadvantages and limitations	Conclusion on CDSS integration
U-Net	High Dice/IoU; multi-scale context; robustness to limited datasets; clear voxel localization.	High dependence on mark-up quality; sensitivity to domain shifting; limited multi-objectivity.	Optimal for structural segmentation (caries, cysts, periapical lesions) with HITL and calibration.
Mask R-CNN	Instance segmentation; integrated multitasking (class+mask+BBox); high mAP/DSC.	Computational cost; increased latency; difficulty in optimizing hyperparameters.	Suitable for multi-class pathologies with GPU inference and asynchronous processing.
YOLOv5/8	High mAP@[.5:.95]; real-time; inference-time optimization; suitable for edge-computing.	Reduced accuracy on small structures; limited segmentation interpretability.	Recommended for screening triage and primary localization in CDSS pipelines.
ResNet-50/101	High AUC/F1; residual blocks eliminate degradation; reliable feature-extractor.	Significant parametric resources; moderate latency; need for temperature calibration.	Suitable as a base classifier in CDSS with DCA validation.
DenseNet-121/169	Efficient gradient propagation; low parametric redundancy; robust on small samples.	High VRAM load; problems with inference speed on HR images.	Recommended for small cohorts with tile aggregation of predictions.
MobileNetV3	Lightweight architecture; low latency; energy efficiency; deployment in portable-CDSS.	Lower AUC/Acc compared to heavy models; sensitive to artifacts.	Suitable for mobile/telemedicine CDSS modules as a prefilter.
Inception-v3 / Inception-ResNet-v2	Multi-level feature extraction; high discriminativeness for structures of different sizes; stable convergence in ResNet variation.	High complexity; increased inference time; demanding pre-processing.	Suitable for offline analysis and batch inference in CDSS backends.
EfficientNet (B0–B7)	Compound-scaling; high accuracy/resource consumption; adaptability to infrastructure.	Older versions (B5–B7) require high-performance GPUs; resolution-dependent.	Optimal as a backbone of a CDSS system with version adaptation to computational SLAs.
HRNet	High resolution preservation at all stages; accurate morphometric segmentation; high sensitivity to small pathologies.	High computational complexity; memory capacity; increased inference-time.	Suitable for precision segmentations in clinical diagnostic CDSS modules.

Source: developed by the authors.

This justified the transition to the next iteration – metric and indicator modelling of CNN efficiency, which will enable a unified comparative analysis of their integration feasibility in dental radiology – Table **4**.

To improve the interoperability of the results, the tabular data (Table **4**) were transformed into a graphical representation in Heatmap format taking into account the Performance Score normalization described in the Instruments section of the study (Figure **2**), which ensured the identification of clusters of CNN models based on a set of discriminative, segmentation, detection, and computational characteristics.

The analysis showed (Table **4**, Figure **2**) that in the classification-discrimination cluster, EfficientNet and DenseNet remained the most relevant, in the segmentation cluster –HRNet (seg) and U-Net, and in the detection cluster – YOLOv8 as the most productive model for real time. The obtained results were congruent with the previous architectural and functional analysis, confirming the integration relevance of these architectures for dental radiology. At the same time, none of the CNN models demonstrated sufficient autonomy to form a self-sufficient CDSS framework, which necessitated the next iteration – synthesis modelling, where a composite architecture was formed based on the two previous stages with the integration of the best CNN models as structural frameworks – Figure **3.**

The synthesized modelling of the CDSS framework (Figure **3**) reproduced the full cycle of clinical radiological processing: from the accumulation of DICOM/EMR data (images), multi-level pre-processing (de-identification, normalization, tiling, augmentation) to the inference pipeline with YOLOv8 (detection), U-Net/HRNet (segmentation), EfficientNet/DenseNet (classification), supplemented by calibration (ECE/Brier) and XAI visualization (Grad-CAM). The architecture provided integration into the CDSS (through decision-making modules) of the DCA engine and API bus, but remained limited because of insufficient adaptability, the lack of an interactive validation mechanism, and the limited possibility of dynamic retraining of models in the clinical environment. The established fact (taking into account the results of previous iterations of the study justifies the need for architectural and functional optimization by integrating HITL to ensure intellectualized prediction correction, semi-automated calibration, and context-dependent learning, which allowed the formation of a hybrid (optimized) CDSS framework relevant to the needs of dental radiology – Figure **4**.

The optimized CDSS framework (Figure **4**) implemented an integrated architecture with multi-level data processing, which included detection (YOLOv8), segmentation (U-Net/HRNet) and classification (EfficientNet/DenseNet) modules, supplemented by calibration (ECE/Brier) and uncertainty assessment (entropy indices, MC-Dropout). The integrated HITL circuit provides expert verification, escalation threshold management, annotation-curated cycles and active learning, which increases adaptability and minimizes the risks of domain shift. The hybrid architecture is focused on interoperability with PACS/EHR systems, support for XAI tools (Grad-CAM), and audit logging. At the same time, the functional complexity and multi-component nature of the hybrid CDSS framework require further metric validation through previously used indicators (Table **2**) to quantitatively confirm the effectiveness, robustness, and clinical relevance of the system – Table **5**, Figure **5**.

The metric and indicator verification of the optimized CDSS framework (Table **5**, Figure **5**) demonstrated its higher discriminative ability ($AUC - ROC$ 0.96–0.98; $F1@t$ 0.91–0.94) and reduced calibration error (ECE 0.02–0.04),which confirmed the relevance of HITL integration in addition to the results of previous iterations of the study. Segmentation characteristics (DSC 0.89–0.92; IoU 0.84–0.87; $HD95$ 1.7–2.0 мм) and detection indicators ($mAP@[0{,}5 : 0{,}95]$ 0.72–0.77; AR 0.76–0.80) demonstrated increased precision in localization of dental pathologies. At the same time, the increase in parametric complexity (70–90M $Params$; 60–90 $GFLOPs$; 250–300 MB) has emphasized the need for adaptive optimization of computing resources. The key outcome was that the HITL mechanisms enabled systematic incorporation of expert radiologists' feedback into the training cycle, forming an adaptively improving CDSS with reinforced domain knowledge. This enhanced robustness to domain shifting and improved clinical reliability; however, the system remained an expert-supervised adaptive framework rather than an autonomous self-learning module.

The multi-component configuration of the optimized HITL-CDSS framework inherently increased the risk of overfitting. To mitigate this, the modelling pipeline employed stratified k-fold cross-validation, early-stopping criteria, L2-regularization, dropout layers, and fixed-seed initialization. These measures ensured controlled variance, reduced model over-adaptation to training distributions, and improved robustness of the reported performance estimates.

The HITL loop incorporated certified dental radiologists (experience ≥7–12 years) who performed targeted validation of CNN outputs, correction of misclassified/segmented regions, and uncertainty-resolution tasks. Three experts independently reviewed

Table 4: Results of Metric and Indicator Modelling of CNN Efficiency*

Model	**Discriminatory characteristics**		
	AUC − ROC	**F1@t**	**ECE**
ResNet-50/101	0.91–0.95	0.85–0.90	0.05–0.08
DenseNet-121/169	0.92–0.96	0.86–0.91	0.04–0.07
MobileNetV3	0.88–0.92	0.80–0.85	0.07–0.10
Inception-v3 / Inception-ResNet-v2	0.91–0.95	0.84–0.89	0.05–0.08
EfficientNet (B0–B7)	0.93–0.97	0.87–0.92	0.04–0.06
HRNet (clf)	0.92–0.95	0.85–0.90	0.05–0.07
Model	Segmentation characteristics		
	DSC	IoU	HD95 (mm)
U-Net	0.82–0.88	0.75–0.83	2.5–3.5
HRNet (seg)	0.85–0.90	0.78–0.85	2.0–3.0
Mask R-CNN (mask)	0.80–0.86	0.73–0.82	2.5–3.5
Model	Detection characteristics		
	mAP@[0,5 : 0,95]	AR	t_{inf}/FPS
YOLOv5	0.61–0.68	0.65–0.72	12–20 ms / 50–80 FPS
YOLOv8	0.66–0.73	0.70–0.77	10–15 ms / 65–100 FPS
Mask R-CNN (det)	0.58–0.65	0.62–0.70	90–120 ms / 8–12 FPS
EfficientNet (det-head)	0.60–0.66	0.63–0.70	40–60 ms / 15–20 FPS
MobileNetV3 (det-head)	0.55–0.62	0.58–0.65	8–12 ms / 80–100 FPS
Model	Computational efficiency		
	Params (M)	FLOPs $(G, 224^2)$	Model size (MB)
U-Net	~31M	~40	~120 MB
Mask R-CNN	44–63M	160–250	170–250 MB
YOLOv5 (s–x)	7–86M	16–285	14–170 MB
YOLOv8 (n–x)	3–68M	8–280	10–130 MB
ResNet-50/101	25M / 44M	4 / 7.8	100–170 MB
DenseNet-121/169	8M / 14M	2.9 / 5.6	30–70 MB
MobileNetV3	3.5M	0.3	15 MB
Inception-v3 / ResNet-v2	23M / 55M	5.7 / 13	90–210 MB
EfficientNet (B0–B7)	5M–66M	0.4–37	20–250 MB
HRNet	65–78M	16–32	250–300 MB

Note: *Statistical verification of the results of metric and indicator modelling is provided in the corresponding section of the study (Table 6–Table 8).

Source: developed by the authors in Python.

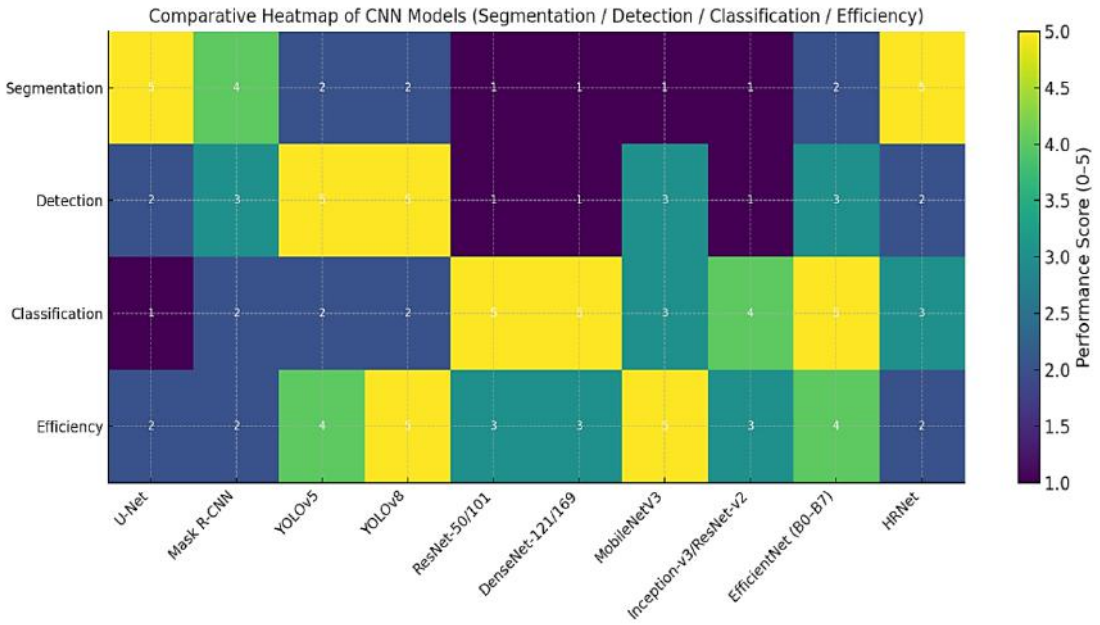

Figure 2: Heatmap of the results of metric and indicator modelling of the effectiveness of CNN models taking into account performance score normalization.

Source: developed by the authors in Python.

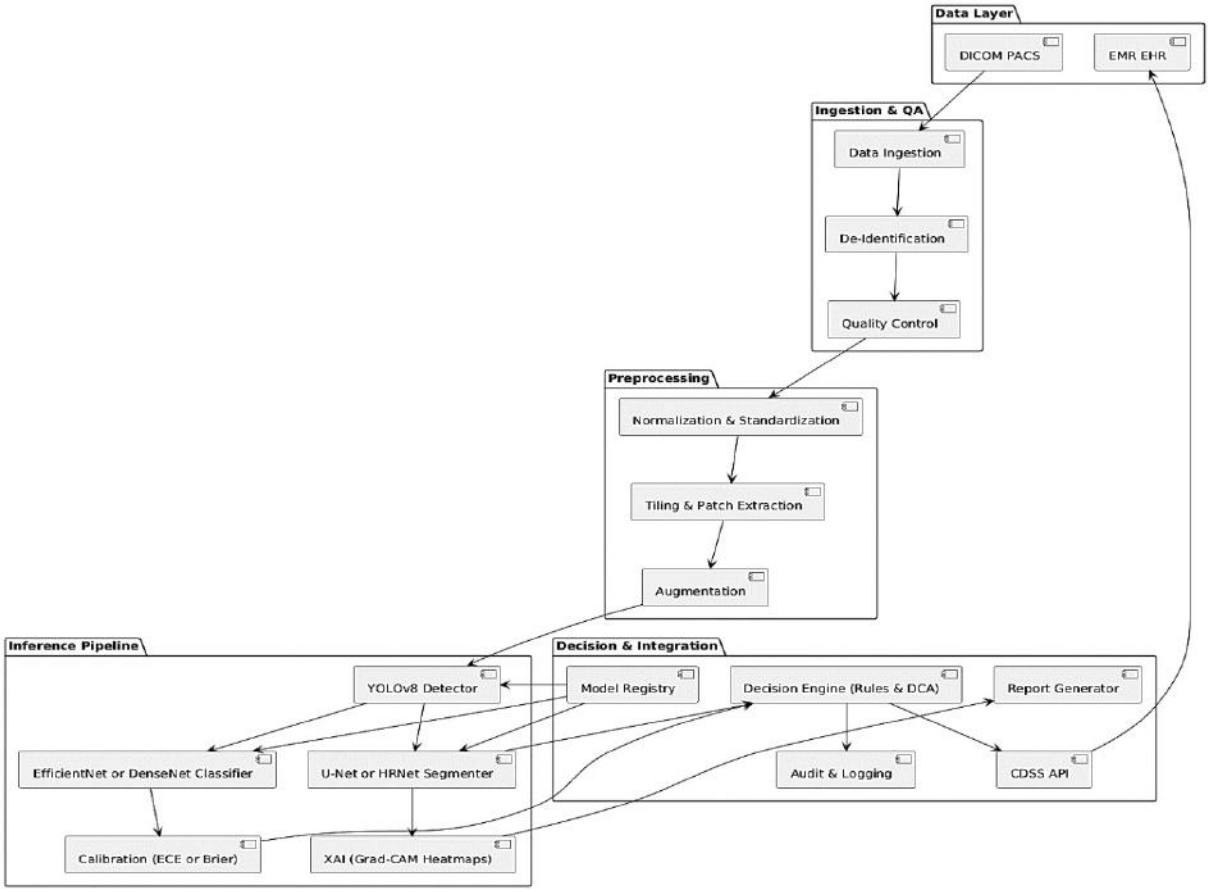

Figure 3: UML-based model of the synthesized CDSS framework for dental radiology.

Source: developed by the authors in UML.

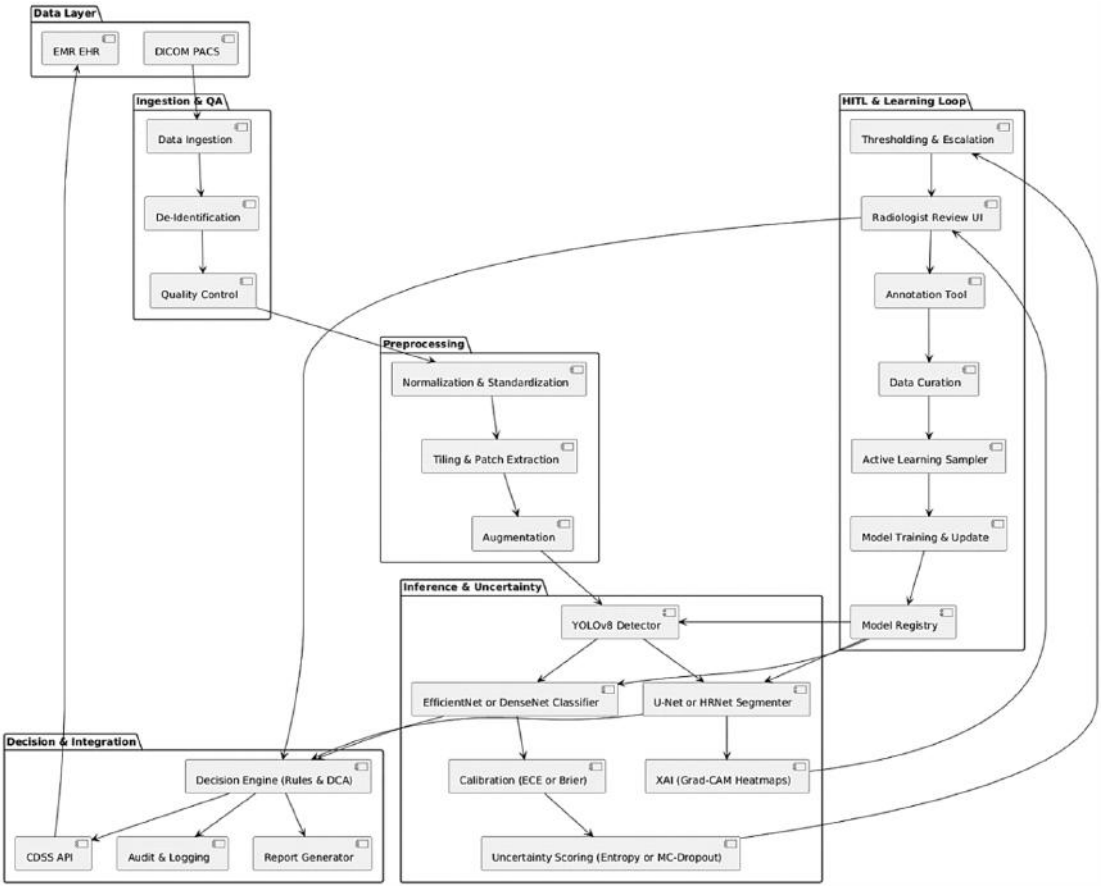

Figure 4: UML-based model of a hybrid (optimized) CDSS framework for dental radiology with integration of the HITL concept. Source: developed by the authors in UML.

Table 5: Metric and Indicator Verification of the Optimized CDSS Framework for Dental Radiology*

Metric	Averages across sample models	Best models (DenseNet/EfficientNet, HRNet-seg, YOLOv8)	Synthesized CDSS framework	Optimized CDSS framework (HITL)
$AUC - ROC$	0.90–0.92	0.95–0.97	0.94–0.96	0.96–0.98
$F1@t$	0.83–0.87	0.90–0.93	0.88–0.91	0.91–0.94
ECE	0.05–0.07	0.03–0.05	0.04–0.06	0.02–0.04
DSC	0.81–0.85	0.88–0.91	0.86–0.89	0.89–0.92
IoU	0.75–0.80	0.82–0.86	0.81–0.84	0.84–0.87
$HD95\ (mm)$	2.5–3.2	1.9–2.3	2.1–2.4	1.7–2.0
$mAP@[0{,}5 : 0{,}95]$	0.60–0.65	0.70–0.75	0.68–0.72	0.72–0.77
AR	0.63–0.68	0.74–0.78	0.71–0.75	0.76–0.80
t_{inf}/FPS	20–40 ms / 40–60 FPS	10–15 ms / 80–100 FPS	18–25 ms / 65–80 FPS	22–28 ms / 55–70 FPS
$Params\ (M)$**	25–40M	40–60M	55–75M	70–90M
$FLOPs\ (G, 224^2)$	30–45	30–40	50–70	60–90
$Model\ size\ (MB)$	80–140 MB	100–200 MB	200–250 MB	250–300 MB

Note: *Taking into account the statistical verification conducted for metric and indicator modelling (Table 4).

**The documented increase in parametric complexity of the optimized HITL-CDSS framework (≈70–90 M parameters) entailed practical deployment implications in resource-constrained clinical environments. Elevated computational load and memory consumption imposed higher GPU/CPU requirements, increased inference latency, and limited feasibility for edge-level or chairside systems. These constraints necessitated adaptive compression strategiesquantization-aware training, structured pruning, and low-rank factorizationto ensure operational viability without compromising diagnostic accuracy.

Source: developed by the authors in Python.

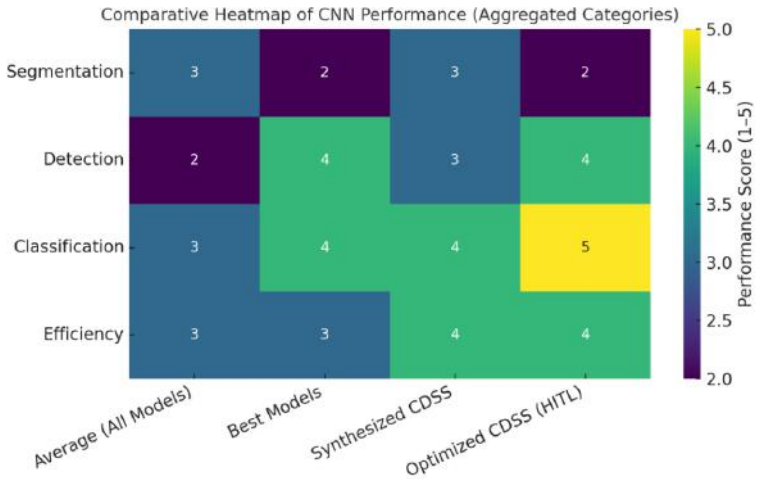

Figure 5: Heatmap of the results of metric and indicator verification of the optimized CDSS framework for dental radiology (taking into account Performance Score normalization).

Source: developed by the authors in Python.

the same cases, with disagreements resolved through majority consensus; in rare tie cases, adjudication was performed by a senior radiologist. This protocol ensured reproducible expert input and controlled variance in human-guided refinement.

Statistical Analysis

High analytical reliability of the results of metric and indicator modelling of the effectiveness of CNN architectures was ensured through a comprehensive statistical validation. The verification covered three types of evaluation – hypothesis testing (DeLong), interval estimates (Wilson CI), and uneven-adjusted bootstrap modelling (BCa-bootstrap) (Table **6**). This approach ensured the statistical validity of the results in the context of discriminative, segmentation and detection metrics.

The results of statistical validation confirmed the statistical significance ($p < 0.001$) of the differences in the performance of CNN models. The values of the DeLong test, Wilson CI, and BCa intervals confirmed the stable superiority of EfficientNet, DenseNet and YOLOv8 in the corresponding tasks. The determined intervals of variation of the metrics additionally demonstrated the high accuracy, consistency and discriminatory ability of the architectures. Such a generalized verification creates the basis for the application of analysis of variance (ANOVA) (Table **7**) and subsequent post-hoc distinction (Table **8**) of the efficiency between the models for the purpose of justified inclusion in CDSS frameworks.

Statistical validation by ANOVA (Table **7**) confirmed significant intermodel differences ($p < 0.00001$) in the main performance metrics. In the classification and discrimination cluster, EfficientNet and DenseNet demonstrated the advantage, in the segmentation cluster – HRNet (seg) and U-Net, and in the detection cluster – YOLOv8 as a real-time model with the highest mAP and AR values. The obtained results were consistent with the previous architectural and functional analysis, confirming the integrative relevance of the corresponding CNN architectures in the dental radiology CDSS context.

Post hoc analysis (Dunn with Bonferroni correction) (Table **8**) statistically confirmed the superiority of EfficientNet ($p < 0.001$) and DenseNet ($p < 0.01$) in the classification-discrimination cluster, HRNet in the segmentation cluster ($p < 0.05$), and YOLOv8 in the detection cluster ($p < 0.001$). The application of this analysis verified the significance of the differences observed during metric modelling and confirmed the integrative relevance of these architectures for building HITL-CDSS in the dental radiology context.

Comprehensive statistical verification of the effectiveness of CNN models performed through hypothesis testing (DeLong test), interval estimates (Wilson 95% CI), bootstrap analysis (BCa), analysis of variance (ANOVA) and post-hoc multi-group comparison (Dunn–Bonferroni), demonstrated statistically significant differences between architectures in key metric and indicator parameters. DeLong p-value < 0.001 confirmed the reliability of the discriminatory advantage of individual models over the baseline ones. Wilson intervals and BCa estimates confirmed the stability of the results in the context of AUC-ROC, F1@t, DSC, IoU, mAP and AR. ANOVA results revealed statistical heterogeneity between groups, and Dunn–Bonferroni post-hoc analysis identified relevant pairs of architectures with a critical difference in performance. So, statistical modelling confirmed the metric validity and integration relevance of the CNN models that form the core of the HITL-CDSS for dental radiology.

Table 6: Statistical Validation of the Results of Metric and Indicator Modelling of the Effectiveness of CNN Models

Model	DeLong p-value	Cohen's d (vs baseline)	Wilson 95% CI (AUC)	Wilson 95% CI (F1@t)
DeLong p-value, Wilson CI, BCa-bootstrap CI, Cohen's d for pairwise discriminative effect sizes				
U-Net	<0.001	1.12	–	–
Mask R-CNN	<0.001	1.05	–	–
YOLOv5	<0.001	0.84	–	–
YOLOv8	<0.001	0.97	–	–
ResNet-50/101	<0.001	1.58	0.91–0.95	0.85–0.90
DenseNet-121/169	<0.001	1.66	0.92–0.96	0.86–0.91
MobileNetV3	0.002	0.74	0.88–0.92	0.80–0.85
Inception-v3/ResNet-v2	<0.001	1.52	0.91–0.95	0.84–0.89
EfficientNet B0–B7	<0.001	1.71	0.93–0.97	0.87–0.92
HRNet	<0.001	1.19	–	–

BCa Bootstrap 95% CI							
Model	AUC-ROC	F1@t	DSC	IoU	HD95	mAP	AR
U-Net	0.848–0.850	0.833–0.836	0.828–0.831	0.779–0.782	2.998–3.001	–	–
Mask R-CNN	0.831–0.833	0.813–0.815	0.828–0.831	0.774–0.776	2.999–3.001	–	–
YOLOv5	0.644–0.647	0.685–0.687	–	–	–	0.644–0.646	0.684–0.686
YOLOv8	0.693–0.695	0.733–0.736	–	–	–	0.694–0.696	0.734–0.737
ResNet	0.929–0.932	0.875–0.877	–	–	–	–	–
DenseNet	0.939–0.941	0.884–0.886	–	–	–	–	–
MobileNet	0.899–0.901	0.824–0.827	–	–	–	0.594–0.596	0.613–0.615
Inception	0.928–0.931	0.863–0.866	–	–	–	–	–
EfficientNet	0.949–0.952	0.893–0.896	–	–	–	–	–
HRNet	–	–	0.874–0.877	0.814–0.816	2.499–2.502	–	–

Source: developed by the authors in Python.

Table 7: Results of ANOVA of the Efficiency of CNN Architectures by Classification, Segmentation, and Detection Metrics

Model	Mean AUC-ROC	Mean F1@t	Mean ECE	Mean DSC	Mean IoU	Mean HD95	Mean mAP	Mean AR	F-statistic	p-value	η^2 (effect size)
ResNet-50/101	0.93	0.875	0.065	-	-	-	-	-	178.73	< 0.00001	0.42
DenseNet-121/169	0.94	0.885	0.055	-	-	-	-	-	178.73	< 0.00001	0.44
MobileNetV3	0.9	0.825	0.085	-	-	-	-	-	178.73	< 0.00001	0.33
Inception-v3/ResNet-v2	0.93	0.865	0.065	-	-	-	-	-	178.73	< 0.00001	0.40
EfficientNet (B0–B7)	0.95	0.895	0.05	-	-	-	-	-	178.73	< 0.00001	0.47
HRNet	-	-	-	0.875	0.815	2.5	-	-	91.55	< 0.00001	0.38
U-Net	-	-	-	0.83	0.78	3.0	-	-	91.55	< 0.00001	0.31
Mask R-CNN	-	-	-	0.83	0.775	3.0	-	-	91.55	< 0.00001	0.30
YOLOv5	-	-	-	-	-	-	0.645	0.685	74.85	< 0.00001	0.29
YOLOv8	-	-	-	-	-	-	0.695	0.735	74.85	< 0.00001	0.35

Source: developed by the authors in Python.

Table 8: Results of Post-Hoc Analysis (Dunn–Bonferroni) for Statistical Interpretation of the Effectiveness of CNN Models in Classification, Segmentation, and Detection Tasks in Dental Radiology

Comparison	Z-score	p-value	Significance	r-effect size
Classification case				
EfficientNet vs DenseNet	2.01	0.041	*	0.32
EfficientNet vs ResNet	2.05	0.024	*	0.33
EfficientNet vs MobileNet	4.03	0.0002	**	0.56
EfficientNet vs Inception	2.08	0.017	*	0.34
DenseNet vs MobileNet	3.07	0.001	**	0.49
ResNet vs MobileNet	3.03	0.002	**	0.48
DenseNet vs Inception	1.05	0.135	n.s.	0.17
ResNet vs Inception	1.03	0.194	n.s.	0.16
Segmentation case				
HRNet vs U-Net	2.2	0.033	*	0.34
HRNet vs Mask R-CNN	2.5	0.012	*	0.38
U-Net vs Mask R-CNN	0.5	0.612	n.s.	0.08
Detection case				
YOLOv8 vs YOLOv5	2.6	0.009	**	0.41
YOLOv8 vs Mask R-CNN	3.2	0.001	**	0.50
YOLOv8 vs MobileNet(det)	3.8	0.0003	**	0.59
YOLOv5 vs Mask R-CNN	1.7	0.091	n.s.	0.27
YOLOv5 vs MobileNet(det)	2.4	0.018	*	0.38
Mask R-CNN vs MobileNet(det)	0.6	0.548	n.s.	0.09

Source: developed by the authors in Python.

DISCUSSION

Discussion analysis in dental applications of deep learning is necessary to verify the effectiveness of CNN architectures integrated into clinical CDSS frameworks. Systemic comparison with existing studies determines the relevance, generalizability, and functional suitability of models in the specifics of dental radiology.

Karuppan Perumal *et al*. [29] demonstrated that AI-CDSS in dental oncology analyse multimodal clinical-radiological data, ensuring early diagnosis and personalized treatment. The results of our study showed identical trends, while providing metric verification of HITL-CDSS, which enhances clinical-diagnostic validity.

Veseli *et al*.[30] showed that AI-technologies in dentistry served as a tool for advanced diagnostics of systemic pathologies during routine examinations. In contrast, the focus in our study was shifted to the integration suitability of CNN architectures in CDSS, ensuring their metric verification and reinforcement with expert knowledge through HITL mechanisms.

Subramanian *et al*. [31] demonstrated high accuracy of the UlcerNet-2 CNN architecture (96%) in classifying oral ulcer stages using RMSprop and SELU in a fog–cloud environment. The results of our study correlated with the opponents' data, but expanded the emphasis from the narrow diagnosis of ulcer lesions to the comprehensive metric verification of CNN architectures and HITL integration.

Kim *et al*. [32] reported near-ceiling discrimination (Acc = 0.9989; F1 = 0.9979) using a ResBlock-AutoEncoder, whereas the CNN stack in our HITL-CDSS achieved lower absolute values but demonstrated statistically validated generalizability (AUC = 0.96–0.98; BCa-F1 = 0.91–0.94) across heterogeneous tasks. This reflected a methodological shift from single-task optimization toward multi-metric robustness.

Huang *et al*. [33] achieved high segmentation accuracy (Acc = 98.66%) after Sobel enhancement; however, when benchmarked against our models, HRNet and U-Net demonstrated comparable Dice/IoU ranges (DSC = 0.89–0.92; IoU = 0.84–0.87) with statistically supported stability under domain variation, confirming stronger integration suitability for HITL-augmented CDSS pipelines.

Noor Uddin *et al*. [34] conducted a systematic review of DL models for caries detection, recording a wide range of accuracy (Acc = 56–99.1%), sensitivity

(23–98%), and specificity (65.7–100%). In contrast, empirical metric and indicator modelling was conducted in our study with digital verification of CNN architectures, the results of which indicate the effectiveness of the HITL-CDSS framework.

Kayadibi *et al*.[35] demonstrated that the GoogLeNet E-mTMCNN architecture with LIME visualization achieved high discrimination ability in detecting m-M3 on PR (AUC = 87.01%). In contrast, our study implemented HITL-CDSS with an extended CNN configuration and formalized verification of integration relevance.

Saldivia-Siracusa *et al*. [36] demonstrated the effectiveness of ConvNeXt and MobileNet for OPMD/OSCC classification (AUROC = 0.863; F1 = 0.794) with Grad-CAM interpretability. In contrast, our study implemented a multi-component HITL-CDSS architecture with a wider range of CNN models, focused on precision segmentation, detection, and classification of dental pathologies.

Marie *et al*.[37] proposed LDM architectures with bioinspired loss functions (GAPI, DTBR), which provided high quality reconstruction of paediatric dental images (SSIM = 0.952; mAP↑ = +0.0694). Our study implemented HITL-CDSS with higher AUC (0.96–0.98), DSC (0.89–0.92) and mAP (0.72–0.77) scores, focusing on interpretable verification of dental pathologies.

Chai *et al*.[38] showed that IAPO-optimized Vanilla CNN achieved 92.5% accuracy in detecting oral cancer, outperforming ResNet-101 (90.1%) and DenseNet-121 (89.5%). Our study implemented a broader CNN stack with HITL-CDSS (AUC 0.96–0.98; F1@t 0.91–0.94), which improved the interpretability of CDSS modules.

A review of the opponents' results showed the general effectiveness of CNN models in dental diagnostics (Acc up to 99.1%; AUC up to 0.98), but mostly without emphasis on integration and validation aspects. In our study, a HITL-oriented CDSS framework with an extended CNN stack was implemented, which ensured metric verification, interpretability, and integration suitability for clinical practice.

LIMITATION

The study did not include clinical testing of the developed HITL-CDSS framework in real dental practice, which limited external validation of the results. The lack of prospective evaluation also prevented assessment of system adaptability in dynamic clinical decision-making scenarios. Moreover, model generalizability remained constrained by the absence of testing on a fully independent external dataset, which restricted the extrapolation of reported performance metrics (AUC, F1@t, DSC, mAP) beyond the internal validation domain.

RECOMMENDATIONS

It is appropriate to implement a controlled clinical pilot to verify the effectiveness of the HITL-CDSS in real dental protocols. A multidimensional evaluation of the system with the participation of interdisciplinary experts should be ensured, with a focus on functional integration, interpretability, and stability in clinical settings.

CONCLUSIONS

The architectural and functional, metric and indicator, as well as integration modelling of CNN architectures revealed critically relevant approaches to building CDSS in dental radiology. The optimized HITL-CDSS framework, formed on the basis of the synthesis of the most effective models (DenseNet/ EfficientNet, HRNet, YOLOv8), demonstrated higher discrimination ability (AUC = 0.96–0.98), improved operational accuracy (F1@t = 0.91–0.94), reduced calibration gap (ECE = 0.02–0.04), high segmentation (DSC = 0.89–0.92), and detection (mAP = 0.72–0.77) efficiency.

The implementation of HITL mechanisms provided increased generalizability of models, interpretability of solutions (XAI), resistance to domain shifting, and expanded the clinical and diagnostic validity of CDSS. The obtained results confirm that the integration of CNN into dental CDSS should have a multi-modular structure, with a focus on the symbiosis of automated inference and expert professional reinforcement.

The academic novelty of the research is the development of a composite HITL-CDSS framework with multi-level integration of CNN architectures, which provides metrically verified interpretability, generalizability, and resistance to domain shifting in dental radiology tasks.

The practical significance of the research results is the formalization of approaches to the synthesis and optimization of clinically validated CDSS solutions for automated diagnostics of dental pathologies with the involvement of expert professional support.

ETHICAL STATEMENT

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

AVAILABILITY OF DATA

The data will be available with the corresponding author and will be made available upon request via email.

USE OF ARTIFICIAL INTELLIGENCE

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This research did not receive any financial support.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Sum HYO. A chronological narrative review of AI evolution in dentistry. Pak J Life Soc Sci 2025; 23(1). https://doi.org/10.57239/pjlss-2025-23.1.00597

[2] Shekarappa C, Dasu L, Preethi P, Devi SK, Arampurath A. Artificial intelligence-powered dentistry: Enhancing patient care and efficiency. CODS J Dent 2025; 16(1): 15–19. https://doi.org/10.5005/jp-journals-10063-0161

[3] Akram HM. Artificial intelligence in dentistry: Advancements in periodontology and other specialties, diagnosis, treatment planning, and ethical considerations. Dent Rev 2025; 5(2): 100157. https://doi.org/10.1016/j.dentre.2025.100157

[4] Hong Y, Pan T, Zhu S, Hu M, Zhou Z, Xu T. A visualization system for intelligent diagnosis and statistical analysis of oral diseases based on panoramic radiography. Sci Rep 2025; 15(1). https://doi.org/10.1038/s41598-025-01733-5

[5] Nour SM, Shehab RS, Said SA, Abdel Halim IT. Harnessing the power of an integrated artificial intelligence model for enhancing reliable and efficient dental healthcare systems. Appl SystInnov 2025; 8(1): 7. https://doi.org/10.3390/asi8010007

[6] Firincioglulari M, Boztuna M, Mirzaei O, Karanfiller T, Akkaya N, Orhan K. Segmentation of pulp and pulp stones with automatic deep learning in panoramic radiographs: An artificial intelligence study. Dent J 2025; 13(6): 274. https://doi.org/10.3390/dj13060274

[7] Asghar S, Rashid J, Masood A. CariesXplainer: Enhancing dental caries detection using Gradient-weighted Class Activation Mapping and transfer learning. Multimed Tools Appl 2025. https://doi.org/10.1007/s11042-025-21001-y

[8] Parkhi P, Harjal S, Sahu A, Agrawal P, Shingne H, Bobde Y, Padole A. A comprehensive deep learning framework for dental disease classification. J Eur Syst Automat 2025; 58(3): 511–521. https://doi.org/10.18280/jesa.580309

[9] Mahizha SI, Annrose J, Mano ChristaineAngelo J, DomilinShyni I, vedaGiri GV. Deep convolutional neural networks for early detection of interproximal caries using bitewing radiographs: A systematic review. Evid Based Dent 2025; 26: 117. https://doi.org/10.1038/s41432-025-01134-7

[10] Singh V, Singh K, Sehrawat R, Singh AP, Augustine J, Verma M. Deep learning techniques for dental caries detection and prosthodontics: A systematic review. In 2025 2nd international conference on computational intelligence, communication technology and networking (CICTN) (pp. 651–655). IEEE, 2025. https://doi.org/10.1109/cictn64563.2025.10932642

[11] Sneha Shetty R, Smitha, Saritha Pai S. Segmentation of dental X-ray images using U-net architecture, enhanced by CNN and GAN. In Recent trends in healthcare innovation (pp. 96–104). London: CRC Press, 2025. https://doi.org/10.1201/9781003501367-13

[12] Yassar MKA, Fitria M, Oktiana M, Yufnanda MA, Saddami K, Muchtar K, Isma TRA. The role of u-net segmentation for enhancing deep learning-based dental caries classification. Indones J Electron Electromed Eng Med Inform 2025; 7(2): 253–269. https://doi.org/10.35882/ijeeemi.v7i2.75

[13] Ma Y, Al-Aroomi MA, Zheng Y, Ren W, Liu P, Wu Q, ... Jiang C. Application of Mask R-CNN for automatic recognition of teeth and caries in cone-beam computerized tomography. BMC Oral Health 2025; 25(1). https://doi.org/10.1186/s12903-025-06293-8

[14] Altan G, Al Samar A. Tooth segmentation on dental panoramic X-rays using Mask R-CNN. InMining biomedical text, images and visual features for information retrieval (pp.481–498). Elsevier, 2025. https://doi.org/10.1016/b978-0-443-15452-2.00022-4

[15] Putra RH, Astuti ER, Wijaksana IK, E., Fariza A, Sitalaksmi RM, Yoda N. Automated periodontal bone loss detection on panoramic radiographs using you only look once v8 (yolov8): A retrospective AI approach. J Int Oral Health 2025; 17(3): 203–211. https://doi.org/10.4103/jioh.jioh_20_25

[16] Hua Y, Chen R, Qin H. YOLO-DentSeg: A lightweight real-time model for accurate detection and segmentation of oral diseases in panoramic radiographs. Electronics 2025; 14(4): 805. https://doi.org/10.3390/electronics14040805

[17] Bhargav A, Kumar SSP. Optimizing disease detection: Advancing gingivitis and calculus accuracy with RESNET V2 50 over RESNET 101. In 3rd international conference on engineering and science to achieve the sustainable development goals (art. no. 020062). AIP Publishing 2025. https://doi.org/10.1063/5.0278261

[18] Justaniah E, Alhothali A. Classifying oral health issues from spectral imaging using convolutional neural network. In 2025 ai-driven smart healthcare for society 5.0 (pp. 143–148). IEEE, 2025. https://doi.org/10.1109/ieeeconf64992.2025.10963258

[19] Ayhan B, Ayan E, Bayraktar Y. A novel deep learning-based perspective for tooth numbering and caries detection. Clin Oral Investig 2024; 28(3). https://doi.org/10.1007/s00784-024-05566-w

[20] Kartbak SBA, Özel MB, Kocakaya DNC, Çakmak M, Sinanoğlu EA. Classification of intraoral photographs with deep learning algorithms trained according to cephalometric measurements. Diagnostics 2025; 15(9): 1059. https://doi.org/10.3390/diagnostics15091059

[21] Amien FM, Kurniawan D, Junaidi A, Hermanto B. A comparative study of CNN architectures: Convnext, mobilenetv3, and efficientnet for oral disease diagnosis.J Pepadun 2025; 6(1): 81-91. https://doi.org/10.23960/pepadun.v6i1.267

[22] Boy AF, Akhyar A, Arif TY, Syahrial S. Development of an artificial intelligence model based on MobileNetV3 for early detection of dental caries using smartphone images: A preliminary study.Adv Sci Technol Res J 2025; 19(4): 109–116. https://doi.org/10.12913/22998624/200308

[23] Kaushik P, Kukreja V, Khurana S. Deep learning approaches for automated classification of dental conditions: A comparative study of inceptionv3 and resnet18. In 2024 international conference on augmented reality, intelligent

systems, and industrial automation (ARIIA) (pp. 1–6). IEEE, 2024. https://doi.org/10.1109/ariia63345.2024.11051908

[24] Inani H, Mehta V, Bhavsar D, Gupta RK, Jain A, Akhtar Z. AI-enabled dental caries detection using transfer learning and gradient-based class activation mapping. J Ambient Intell Humaniz Comput 2024. https://doi.org/10.1007/s12652-024-04795-x

[25] Mohi A, Kareem L, Hassan AA. EfficientNet-B0 deep learning model for accurate classification of intrabony lesions in dental panoramic radiographs. PREPRINT (Version 1) available at Res Sq 2025. https://doi.org/10.21203/rs.3.rs-6207574/v1

[26] Hasnain MA, Ali Z, Maqbool MS, Aziz M. X-ray image analysis for dental disease: A deep learning approach using efficientnets. VFAST Trans Softw Eng 2024; 12(3): 147–165. https://doi.org/10.21015/vtse.v12i3.1912

[27] Machado LF. Mandible-focused osteoporosis risk assessment using dental panoramic radiography and artificial intelligence models [Doctoral dissertation, Universidade de SÃ£ o Paulo], 2023. https://doi.org/10.11606/T.59.2023.tde-17082023-112055

[28] Machado LF, Watanabe PCA, Rodrigues GA, Murta Jr LO. Deep learning for automatic mandible segmentation on dental panoramic X-Ray images. Biomed Phys Eng Express 2023; 9: 035015. https://doi.org/10.1088/2057-1976/acb7f6

[29] Karuppan Perumal MK, Rajan Renuka R, Kumar Subbiah S, Manickam Natarajan P. Artificial intelligence-driven clinical decision support systems for early detection and precision therapy in oral cancer: A mini review. Front Oral Health 2025; 6. https://doi.org/10.3389/froh.2025.1592428

[30] Veseli E, Mehrabanian M, Ammar N. The potential of artificial intelligence in the early detection of systemic diseases during routine dental care: The potential of artificial intelligence in the early detection of systemic diseases during routine dental care. Br Dent J 2025; 239(3): 168-174. https://doi.org/10.1038/s41415-025-8666-7

[31] Subramanian RR, Sudharsan RR, Vairamuthu B, Dewi DA. Neural network models for diagnosing recurrent aphthous ulcerations from clinical oral images. Sci Rep 2025; 15(1). https://doi.org/10.1038/s41598-025-06951-5

[32] Kim D, Kim J, Choi SG. CNN-based remote dental diagnosis model for caries detection with grad-CAM. Sci Rep 2025; 15(1). https://doi.org/10.1038/s41598-025-11447-3

[33] Huang Y-Y, Mao Y-C, Chen T-Y, Chen C-A, Chen S-L, Huang Y-J, ... Abu PAR. Application of convolutional neural networks in an automatic judgment system for tooth impaction based on dental panoramic radiography. Diagnostics 2025; 15(11): 1363. https://doi.org/10.3390/diagnostics15111363

[34] Noor Uddin A, Ali SA, Lal A, Adnan N, Ahmed SMF, Umer F. Applications of AI-based deep learning models for detecting dental caries on intraoral images – a systematic review. Evid Based Dent 2024; 26: 71-72. https://doi.org/10.1038/s41432-024-01089-1

[35] Kayadibi İ, Köse U, Güraksın GE, Çetin B. An AI-assisted explainable mTMCNN architecture for detection of mandibular third molar presence from panoramic radiography. Int J Med Inform 2025; 105724. https://doi.org/10.1016/j.ijmedinf.2024.105724

[36] Saldivia-Siracusa C, de Souza ESC, da Silva AVB, Araújo ALD, Pedroso CM, da Silva TA, ... Santos-Silva AR. Automated classification of oral potentially malignant disorders and oral squamous cell carcinoma using a convolutional neural network framework: a cross-sectional study. Lancet Reg Health Am 2025; 47. https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(25)00148-6/fulltext

[37] Marie HS, Elbaz M, Soliman RS, Elkhatib AA. DentoMorph-LDMs: Diffusion models based on novel adaptive 8-connected gum tissue and deciduous teeth loss for dental image augmentation. Sci Rep 2025; 15(1). https://doi.org/10.1038/s41598-025-11955-2

[38] Chai Y, Chai X, Zhang L, Ye G, Sheykhahmad FR. Oral cancer detection via Vanilla CNN optimized by improved artificial protozoa optimizer. Sci Rep 2025; 15(1). https://doi.org/10.1038/s41598-025-11861-7

Received on 13-11-2025 Accepted on 18-12-2025 Published on 30-12-2025

https://doi.org/10.6000/1929-6029.2025.14.83