Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Safety Envelope for Foundation-Model Dental Diagnosis: Bounded-Deviation Guarantees and Cross-Center Refusal

Prof. Ayman Elnashar

ayman.elnashar@ku.ac.ae

Khalifa University https://orcid.org/0000-0001-7986-4474

Dr. Norsin Elnashar

Ajman University

Prof. Ahmed Elemam

King Khalid University

Research Article

Keywords: Dental radiography, foundation models, calibration, out-of-distribution detection, trustworthy machine learning, safety guarantees, panoramic X-ray, selective prediction

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10255012/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Safety Envelope for Foundation-Model Dental Diagnosis: Bounded-Deviation Guarantees and Cross-Center Refusal

A. Elnashar, N. Elnashar, A. Emam

***Abstract*—Foundation models for dental radiograph analysis have advanced rapidly on diagnostic accuracy, yet none provide the reliability properties a clinical decision-support tool requires calibrated confidence, a bounded and recoverable relationship to a trusted baseline, and the ability to abstain on inputs from an unfamiliar source. We introduce a backbone-agnostic *safety envelope* that wraps any frozen foundation encoder and supplies these properties without retraining the backbone. A temperature-scaled linear probe defines a calibrated baseline; a shallow residual head adds a per-condition correction hard-clipped to a bound δ, so the final prediction provably deviates from the baseline by at most δ and recovers it exactly in the worst case. An out-of-distribution gate based on embedding-space Mahalanobis distance abstains on inputs far from the training distribution. We first establish, with a three-architecture convolutional zoo and a multi-label head ablation, that accurate dental classifiers are routinely miscalibrated—the gap the envelope is designed to close. On the DENTEX panoramic benchmark, under common-random-number replication across three seeds and two structurally different encoders—a dental ViT-L and a biomedical CLIP—the envelope improves calibration where the backbone needs it (expected calibration error 0.040→0.034 on the dental encoder) while preserving discrimination (macro-AUC 0.849→0.857) and holding the deviation guarantee in all eighteen tested conditions. On the Tufts Dental Database, a genuinely different clinical center, the envelope refuses 85% of inputs at an operating point that refuses only 5% of its own center. The method is one small head and one distance computation on top of a frozen encoder; it is agnostic to the backbone and drops in directly as stronger dental foundation models are released.**

***Index Terms*—Dental radiography, foundation models, calibration, out-of-distribution detection, trustworthy machine learning, safety guarantees, panoramic X-ray, selective prediction.**

I. Introduction

Panoramic dental radiography is among the most frequently acquired diagnostic images in medicine, and automated abnormality analysis on panoramic films has become an active research frontier. The last two years have produced a rapid sequence of increasingly capable systems. Notation-aware vision–language detectors exploit the fixed topology of the dental arch and standardized tooth numbering to localize pathology [1]; dental-specialized multimodal large language models reason over panoramic films and surpass general-purpose models on dental question answering and report generation [2]; and self-supervised dental vision foundation models, pretrained on the order of a million unlabeled radiographs, supply transferable features across detection, segmentation, and landmarking [3]. Almost without exception, these systems are compared on a single axis—detection or classification accuracy.

Accuracy on a held-out test set, however, is not the same as sufficiency for clinical deployment. A model that is accurate on average can still be confidently wrong on an individual film. It can be *miscalibrated*, so that a reported probability of 0.6 does not correspond to a 60% empirical frequency, rendering its confidence scores unusable for triage or thresholding. It can *degrade silently* relative to a trusted reference when redeployed, with no contract bounding how far its outputs may drift. And it can *fail without warning* when presented with an image from a clinic, scanner, or population it has never encountered, asserting a fluent but unfounded prediction. None of these failure modes is addressed by improving mean accuracy, and none of the recent dental foundation systems exposes a mechanism to detect or bound them. The gap is not accuracy; it is *safety*.

This paper contributes a layer that sits on top of an accurate model and supplies the missing safety properties. The design follows three principles. First, predictions must be **calibrated**, so that stated confidence is meaningful. Second, the deployed model must deviate from a trusted, calibrated baseline by only a **bounded** amount and must recover that baseline in the worst case—a contract that structurally precludes silent degradation. Third, the model must be able to **refuse**, deferring to a clinician when an input lies outside its competence rather than asserting an unreliable prediction, in the spirit of selective classification [4]. Together these constitute a *safety triad*: bounded deviation, calibration, and refusal.

We realize the triad as a backbone-agnostic *safety envelope* operating on frozen foundation-model embeddings. The frozen-backbone setting is deliberate: it is where a safety layer is most portable, since the identical envelope transfers across encoders without retraining. Our contributions are:

(C1) a residual envelope head whose hard clip enforces a provable bound δ on the deviation from a calibrated baseline, with exact recovery of that baseline at the worst case (Proposition 1);

(C2) an embedding-space out-of-distribution gate, based on Mahalanobis distance, that abstains on unfamiliar inputs at a threshold tuned to a chosen in-distribution refusal rate;

(C3) evidence, from a three-architecture classifier zoo and a multi-label head ablation, that accurate dental models are systematically miscalibrated—motivating the envelope—together with a demonstration, under common-random-number replication across two structurally different encoders, that the

envelope improves calibration where the backbone needs it while never harming discrimination; and

(C4) a cross-center evaluation on a genuinely different clinical database showing the refusal mechanism defers the large majority of unseen-center inputs at an operating point tuned to its own center. The complete method adds one small head and one distance computation to a frozen encoder and is immediately reusable as stronger dental backbones appear.

II. Related Work

A. Dental detectors and foundation models

Recent dental systems pursue accuracy through scale and domain specialization. Notation-aware detectors prompt a vision–language model with standardized tooth numbering and the left–right symmetry of the arch, improving mean average precision over generic object detectors on panoramic abnormality detection. [1] Dental multimodal large language models add structured reasoning over panoramic films and outperform general-purpose models on dental benchmarks and report generation. [2] Self-supervised dental vision foundation models, trained on roughly a million unlabeled radiographs, yield features that transfer across diagnosis, segmentation, and landmark localization. [3] Early dental deep-learning systems established strong single-task performance on caries and bone-loss detection [5], [6], [7], [8], with later work extending to caries and lesion segmentation [9] and to robust segmentation under noisy labels [10] and to integrated dental health-informatics platforms [11]. These advances are genuine and complementary to ours. Crucially, none equips its output with calibration, a bounded-deviation contract, or principled refusal; our envelope is designed to wrap exactly such systems and supply those properties post hoc.

B. Calibration and selective prediction

Temperature scaling is a standard single-parameter post-hoc calibration method fit on a held-out split, known to reduce expected calibration error without altering rank order. [12], [13] Selective prediction and out-of-distribution detection allow a model to abstain [14], [4]; among detectors, the Mahalanobis distance to the class-conditional or marginal feature distribution is a widely used, lightweight score. [14], [15] Our contribution is not a new calibrator or detector in isolation, but their integration into a single envelope whose central element—the bounded, recoverable residual—supplies a formal worst-case guarantee that calibration and refusal alone do not provide. Complementary uncertainty approaches—deep ensembles [16], Monte-Carlo dropout [17], evaluation under dataset shift [18], the ODIN detector [19], post-hoc calibration under drift [20], revisited modern-network calibration [13], and surveys of OOD detection in medical imaging [21]—quantify uncertainty but do not impose a hard, per-input deviation bound.

C. Frozen foundation embeddings with lightweight heads

A growing body of work attaches lightweight probes or adapters to frozen foundation embeddings rather than fine-tuning the backbone [22], [23], [24], trading a small accuracy cost for label efficiency, stability, and reuse. We adopt this setting because it is the one in which a portable safety layer is most valuable: the same envelope, with the same guarantee, transfers unchanged across encoders, as we demonstrate on a dental transformer and a biomedical vision–language model.

III. Method

A. Problem setup and notation

We address multi-label diagnosis at the level of an individual tooth. Each tooth is represented by a crop from a panoramic radiograph with a label vector y over K diagnostic conditions, where a tooth may carry several conditions simultaneously. A frozen foundation encoder φ maps a crop to an embedding,

$$\boldsymbol{x} = \varphi(\text{crop}) \in \mathbb{R}^d \tag{1}$$

with embedding dimension d, obtained from a vision-transformer encoder [25]. Throughout, vectors are set in lowercase bold italic and matrices in uppercase bold italic, scalars in italic, and operators, numerals, and function names upright. All learning occurs on top of x; the encoder is never updated, so its embeddings are computed once and cached.

B. Calibrated baseline

For each condition k an ℓ_2-regularized logistic probe on the training embeddings yields a logit $z_k(x)$. A single temperature T is fit on the validation split by minimizing binary cross-entropy of the temperature-scaled logits [12],

$$T^* = \arg\min_T \sum_k \text{BCE}\left(\frac{z_k}{T}, \boldsymbol{y}_k\right) \tag{2}$$

producing the calibrated baseline

$$b_k(\boldsymbol{x}) = \frac{z_k(\boldsymbol{x})}{T}, \quad p_k^{\text{base}}(\boldsymbol{x}) = \sigma\big(b_k(\boldsymbol{x})\big) \tag{3}$$

where σ is the logistic function. The baseline is intentionally simple and trustworthy; the envelope's guarantee is stated relative to it.

C. Bounded-deviation safety envelope

A shallow residual head $r(x)$ produces a per-condition correction hard-clipped to the symmetric bound δ and added to the baseline logit,

$$f_k(\boldsymbol{x}) = b_k(\boldsymbol{x}) + \text{clip}(\boldsymbol{r}_k(\boldsymbol{x}), -\delta, +\delta) \tag{4}$$

with envelope probability

$$p_k^{\text{env}}(\boldsymbol{x}) = \sigma\big(f_k(\boldsymbol{x})\big) \tag{5}$$

Only the residual head r is trained, by binary cross-entropy against the labels, with the baseline held fixed. The clip is the load-bearing element: it converts an unconstrained correction into a bounded one, echoing worst-case-constrained formulations used elsewhere in robust estimation [26].

D. Recovery guarantee

Proposition 1 (Bounded deviation and recovery). For every input x and condition k,

$$|f_k(\boldsymbol{x}) - b_k(\boldsymbol{x})| \le \delta \tag{6}$$

Consequently, the envelope probability lies within a bounded band of the calibrated baseline probability, and if the residual head is identically zero the envelope recovers the baseline exactly.

Proof. By construction the envelope logit minus the baseline

logit equals the clipped residual, whose magnitude is at most δ by definition of the clip; setting the residual head to zero gives equality of the two logits, hence of the two probabilities.

The guarantee is *structural*, not statistical: it holds for any input, any trained or untrained head, and any backbone, because it is a property of the clip operation rather than of the data. We state it as a design guarantee rather than a deep theoretical result—the bound follows directly from the clip and is intended to be simple and auditable, not surprising. It gives a worst-case bound—the deployed model can never drift further than δ from the trusted baseline—while still permitting useful correction within the band; unlike distribution-free conformal guarantees [27], [28], [29], [30], the bound here is structural and holds for every input rather than in expectation over a calibration draw. The bound δ is the single tunable dial trading correction capacity against deviation risk; its sensitivity is studied in Section V-E.

E. Out-of-distribution refusal

We fit a Gaussian to the training embeddings and score each test embedding by its Mahalanobis distance to that distribution [15],

$$m(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \quad (7)$$

with mean vector μ and covariance matrix Σ. A refusal threshold τ is set at the 95th percentile of the distance on the model's own validation distribution—accepting a 5% refusal rate on familiar data—and any input exceeding τ is deferred to a clinician. The same score yields a refusal–accuracy curve in-distribution and a cross-center refusal rate out-of-distribution. Fig. 1 summarizes the complete system: a frozen encoder feeds a calibrated baseline, a bounded envelope head, and an OOD gate, whose outputs are a calibrated bounded prediction and an accept/refuse decision.

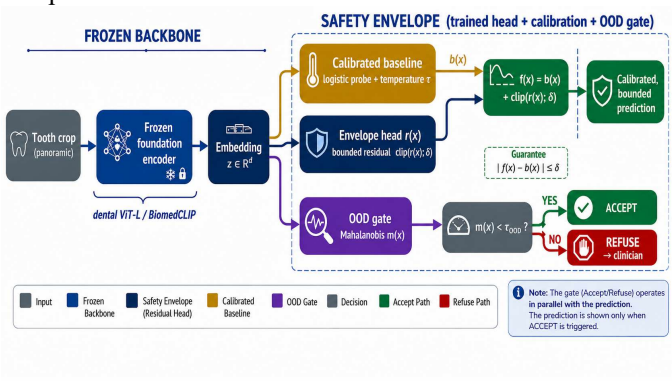


Fig. 1. System overview. A frozen foundation encoder maps a tooth crop to an embedding that feeds three components of the safety envelope: a temperature-scaled calibrated baseline, a residual envelope head whose correction is hard-clipped to ±δ (so |f−b|≤δ by construction), and a Mahalanobis out-of-distribution gate that accepts or refuses. The backbone is never retrained.

IV. Experimental Setup

A. Datasets

DENTEX [31] (primary, in-distribution) provides 705 panoramic radiographs with tooth-level diagnosis annotations under a documented hierarchical protocol. Grouping annotations by tooth yields 3,456 teeth over four conditions: Impacted (604 positives), Caries (2,186), Periapical Lesion (158), and Deep Caries (575). The strong class imbalance—Periapical Lesion is rarest and Caries most common—is representative of clinical prevalence and is reported throughout. **Tufts** [32] (cross-center, out-of-distribution) provides 1,000 labeled panoramic radiographs from a different country and acquisition system, with tooth bounding boxes that we use to extract matched tooth crops for an apples-to-apples comparison. The two databases differ in scanner, population, and acquisition protocol and so constitute a genuine center shift. ADCD [44] (cross-center, out-of-distribution) is a panoramic radiograph collection from which we use the external-validation split (180 images from other clinics, a genuine multi-clinic external set); its YOLO annotations mark three of our four conditions—Caries, Impacted, and Periapical Lesion, with no Deep Caries class—yielding 513 condition crops from 136 images. Because ADCD annotates only pathological findings, it contains no labeled normal teeth; we therefore use it for cross-center refusal behavior, not for accepted-case reliability. Tufts and ADCD span different degrees of shift relative to DENTEX, which we exploit in Section V-G.

B. Backbones

We evaluate two structurally different frozen encoders to probe backbone-agnosticism. **Dental ViT-L** is a ViT-Large transformer from a dental visual-expert family, pretrained on periapical radiographs [2]; we expose its 1,024-dimensional class-token embedding at 512×512 input resolution. Because it was trained on periapical images and is applied here to panoramic crops, it operates under a real periapical→panoramic domain shift, which we treat as an honest test condition rather than conceal. **BiomedCLIP** is a biomedical vision–language encoder (ViT-B, 512-dimensional image embedding) trained on biomedical figure–caption pairs [33]; it is biomedical but not dental, and serves as the agnosticism control. For the domain-shift check reported in Section VII, we additionally employ DINOv3-L, a self-supervised vision transformer (ViT-L, 1,024-dimensional embedding) pretrained on natural images and applied directly to panoramic crops, as a panoramic-capable comparison encoder.

C. Protocol and metrics

Splits are grouped by source image so that all teeth from one radiograph fall entirely within train, validation, or test, preventing patient-level leakage; the seed-42 partition is 2,416 / 511 / 529 teeth. We adopt a common-random-number protocol in which each seed produces an identical partition across all conditions, isolating the experimental manipulation as the only varying factor. Embeddings are extracted once per backbone and cached; the seed loop and δ-sweep re-split and re-train only the lightweight head. Unless noted, results are mean ± standard deviation over seeds {42, 43, 44} at δ = 1.5. We report per-condition and macro-averaged area under the ROC curve (AUC) for discrimination, expected calibration error (ECE) with ten bins for calibration, the maximum observed logit

deviation against δ for the guarantee, and refusal rate and refusal–accuracy curves for the gate. All experiments run locally on an Apple-silicon workstation using the Metal backend.

V. Results

A. Accurate dental classifiers are miscalibrated

We first establish the discrimination achievable by fully task-trained convolutional baselines and the calibration problem that accompanies it. Table I reports a three-architecture zoo trained end-to-end on DENTEX tooth crops. [34], [35], [36] The three models cluster within three AUC points—EfficientNetV2-S 0.919, DenseNet-121 0.913, ResNet-50 0.888—confirming that, once data and protocol are fixed, architecture choice is largely Pareto-indifferent. Per-condition AUC tracks class frequency and clinical difficulty: Impacted is easiest and Deep Caries, definitionally adjacent to Caries, is hardest.

TABLE I
DIRECTION-1 CLASSIFIER ZOO ON DENTEX (TEST MACRO- AND PER-CONDITION AUC, SEED 42). THE TIGHT CLUSTERING ACROSS ARCHITECTURES INDICATES PARETO-INDIFFERENCE; DEEP CARIES IS CONSISTENTLY THE HARDEST CONDITION.

Model	Macro	Impact.	Caries	Periap.	Deep C.
EfficientNetV2-S	0.919	0.99	0.91	0.94	0.84
DenseNet-121	0.913	0.99	0.90	0.94	0.82
ResNet-50	0.888	0.98	0.88	0.88	0.81

Discrimination, however, masks a calibration problem. Fig. 2 shows reliability diagrams for the strongest model. Impacted (ECE 0.017) and Periapical Lesion (0.025) are well calibrated, but Caries (0.092) and especially Deep Caries (0.127) are not: the Deep-Caries curve swings widely about the diagonal, so its stated probabilities are untrustworthy despite acceptable discrimination. This is precisely the failure a safety layer must address, and it is consistent with calibration deficiencies reported across deep dental and medical classifiers [13], [37], [38], [39].

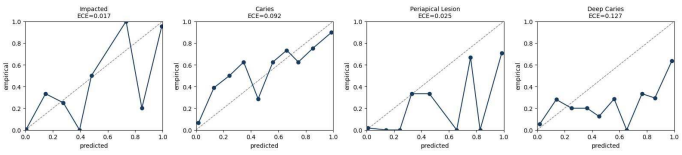


Fig. 2. Reliability diagrams for the strongest task-trained baseline on DENTEX. Discrimination does not imply calibration: Deep Caries (rightmost, ECE 0.127) is badly miscalibrated despite acceptable AUC, motivating the safety envelope.

We also tested whether a more expressive multi-label head closes the calibration gap. Table II ablates a plain sigmoid head against an independent-per-class head and a gated multilayer perceptron. The plain head is Pareto-optimal (macro-AUC 0.9145), with the gated head slightly behind (0.9035): with only four conditions, the gating mechanism has little co-occurrence structure to exploit and adds parameters without payoff. This honest negative result shows that architectural complexity in the head is not the answer—motivating a post-hoc safety layer instead.

TABLE II
DIRECTION-2 MULTI-LABEL HEAD ABLATION (TEST AUC, SEED 42). THE PLAIN HEAD IS PARETO-OPTIMAL; GATING DOES NOT HELP AT FOUR CONDITIONS—AN HONEST NEGATIVE RESULT THAT MOTIVATES A POST-HOC SAFETY LAYER OVER A MORE COMPLEX HEAD.

Multi-label head	Macro-AUC	Periapical	Deep Caries
Plain sigmoid	0.9145	0.914	0.842
Independent per-class	0.9051	0.872	0.845
Gated MLP	0.9035	0.874	0.838

B. Backbone-agnostic improvement under CRN replication

Table III reports the envelope on frozen embeddings across three seeds. On the dental backbone the envelope improves calibration consistently (ECE 0.040→0.034) while slightly improving discrimination (AUC 0.849→0.857); the deviation guarantee holds in every seed. On the biomedical backbone—already well calibrated—the envelope preserves calibration (0.036→0.035) and improves discrimination (0.790→0.813) without harm. The pattern is the intended one: the envelope helps most where the backbone is weakest, and never degrades a backbone that is already strong. The per-seed values confirm the envelope AUC is at or above the baseline in all six runs. We also compared the envelope's temperature-scaled calibration against per-condition isotonic regression and Platt scaling; the per-condition methods can reach a lower macro calibration error, but they fit each class independently and overfit the rarest one—isotonic, for instance, worsens Periapical Lesion calibration relative to the uncalibrated probe. We retain temperature scaling because it is monotonic (it preserves the ranking and hence discrimination), single-parameter, and robust on rare classes, properties that matter more than minimal error for a component sitting beneath a bounded, ranking-preserving guarantee.

TABLE III
ENVELOPE VS. CALIBRATED BASELINE, MEAN ± STD OVER SEEDS {42,43,44} AT Δ = 1.5. THE DEVIATION GUARANTEE HELD IN ALL EIGHTEEN RUNS (THREE SEEDS × TWO BACKBONES × Δ-SWEEP).

Backbone	Base AUC	Env AUC	Base ECE	Env ECE
Dental ViT-L	0.849 ± 0.005	0.857 ± 0.006	0.040 ± 0.002	0.034 ± 0.006
BiomedCLIP	0.790 ± 0.023	0.813 ± 0.014	0.036 ± 0.006	0.035 ± 0.003

To quantify the reliability of these gains we computed bootstrap 95% confidence intervals (2,000 resamples of the test set) on the envelope-minus-baseline macro-AUC. On the dental backbone the gain is statistically resolved (+0.012, 95% CI [+0.005, +0.019], excluding zero); on the biomedical backbone it is not (+0.009, 95% CI [-0.008, +0.026]). We therefore do not rest the contribution on discrimination: the envelope's purpose is calibration, a bounded-deviation contract, and refusal, and the

AUC is reported only to confirm that these are obtained without sacrificing discrimination.

C. Clinical operating-point metrics

Because area under the curve is threshold-independent, we also report standard operating-point metrics at a 0.5 decision threshold (Table IV), before and after the envelope, on the dental backbone. The envelope raises macro-F1 from 0.46 to 0.58, driven mainly by recovered sensitivity on the two conditions the baseline most often misses: deep caries (sensitivity 0.04 to 0.22) and periapical lesion (0.00 to 0.12). Caries sensitivity decreases slightly (0.93 to 0.92) while its specificity rises (0.62 to 0.71), so the net effect is fewer false alarms with a small, explicit trade in caries recall. The very low baseline sensitivity on the two rarest conditions reflects class prevalence rather than a defect: in the test fold, Periapical Lesion has only 25 positives against 504 negatives and Deep Caries 90 against 439, so a baseline probe at a 0.5 threshold almost never crosses it for them and defaults to predicting absence (it recovers 0 of 25 periapical lesions and 4 of 90 deep caries). The envelope's residual head recovers part of this lost recall (1 of 25 and 16 of 90, respectively). Because the positive counts for these conditions are small, their per-condition figures should be read as indicative of direction rather than as precise operating points, and a larger labeled cohort would tighten them. Reducing missed pathology is clinically important, particularly for screening and decision-support, though it must be weighed against the false positives a more sensitive operating point introduces; the envelope concentrates its gain on the former while the gate and calibrated confidence are what bound the latter.

TABLE IV

OPERATING-POINT METRICS AT THRESHOLD 0.5 ON THE DENTAL BACKBONE, BASELINE (B) VS ENVELOPE (E). THE ENVELOPE RECOVERS SENSITIVITY ON THE MOST-MISSED CONDITIONS; MACRO-F1 RISES 0.46 TO 0.58.

Condition	Sens (b/e)	Spec (b/e)	PPV (b/e)	NPV (b/e)	F1 (b/e)
Impacted	0.88/0.90	0.98/0.98	0.92/0.93	0.97/0.98	0.90/0.91
Caries	0.93/0.92	0.62/0.71	0.80/0.84	0.84/0.85	0.86/0.88
Periapical Lesion	0.00/0.12	1.00/0.99	—/0.43	0.95/0.96	0.00/0.19
Deep Caries	0.04/0.22	1.00/0.97	1.00/0.59	0.84/0.86	0.09/0.32
Macro	—	—	—	—	0.46/0.58

D. The guarantee is tight and always satisfied

Table V records the maximum observed logit deviation against δ across the full δ-sweep. In every condition the maximum deviation equals δ to numerical precision: the bound is both never violated and tight (the clip is active), confirming Proposition 1 empirically across eighteen independent runs.

TABLE V

GUARANTEE VERIFICATION ACROSS THE Δ-SWEEP (DENTAL BACKBONE, REPRESENTATIVE). THE MAXIMUM LOGIT DEVIATION EQUALS Δ TO NUMERICAL PRECISION IN EVERY CASE: THE BOUND IS TIGHT AND NEVER VIOLATED.

δ	0.3	0.5	1.0	1.5	2.0	3.0
max |f−b|	0.300	0.500	1.000	1.500	2.000	3.000
holds	✓	✓	✓	✓	✓	✓

E. Sensitivity to the deviation bound

Fig. 3 sweeps δ from 0.3 to 3.0. The dental envelope's AUC stays in a tight band (0.862–0.867), always above the baseline, and its ECE is minimized in the broad region δ ∈ [1.0, 2.0]; the biomedical envelope's calibration improves steadily with δ. The method is robust to its single hyperparameter—there is a wide safe operating region rather than a brittle optimum—and we adopt δ = 1.5 throughout. Table VI lists exact values for the dental backbone.

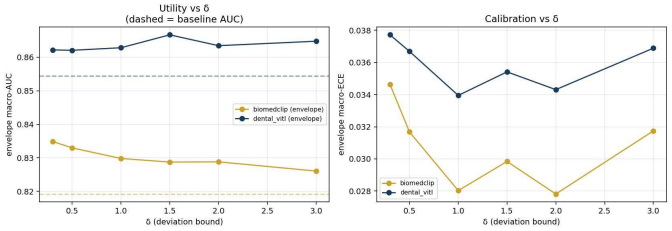


Fig. 3. Safety/utility sensitivity to the deviation bound δ. Left: envelope macro-AUC vs. δ (dashed lines mark the baselines); right: envelope macro-ECE vs. δ. The method is stable across a broad δ region, with δ ∈ [1.0, 2.0] near-optimal for both backbones.

TABLE VI

DENTAL-BACKBONE ENVELOPE ACROSS THE Δ-SWEEP (SEED 42). AUC EXCEEDS THE 0.854 BASELINE AT EVERY Δ; ECE IS BELOW THE 0.041 BASELINE THROUGHOUT, BEST NEAR Δ ∈ [1.0, 2.0].

δ	0.3	0.5	1.0	1.5	2.0	3.0
Env AUC	0.862	0.862	0.863	0.867	0.863	0.865
Env ECE	0.038	0.037	0.034	0.035	0.034	0.037

F. In-distribution refusal

Fig. 4 shows refusal–accuracy curves on held-out DENTEX. On the dental backbone, accuracy on the retained set rises monotonically as the most uncertain inputs are refused (0.896 at full retention to 0.937 when the least-confident tail is deferred), confirming the Mahalanobis score identifies genuinely harder cases. The biomedical curve is flatter, reflecting a less structured embedding geometry—an honest indication that the refusal mechanism is only as informative as the embedding space beneath it.

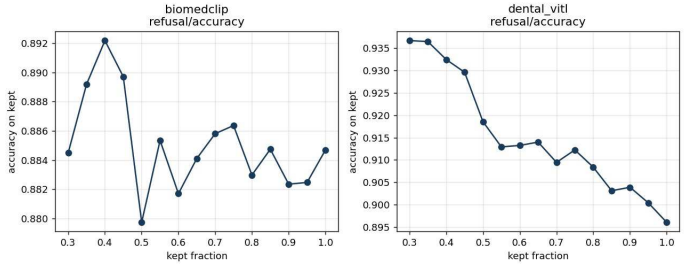


Fig. 4. In-distribution refusal–accuracy curves. Deferring the least-confident tail raises accuracy on the retained dental-backbone set (right), confirming the distance score captures real difficulty; the biomedical curve (left) is flatter.

G. Cross-center refusal on Tufts

The decisive safety test is whether the envelope abstains on a genuinely different center. Fig. 5 plots the Mahalanobis distance of held-out DENTEX crops against Tufts crops, with the refusal threshold fixed at the 95th percentile of DENTEX. The Tufts distribution sits clearly to the right of DENTEX on both backbones. At an operating point refusing only 5.1% of its

own center, the envelope refuses 84.7% of Tufts on the dental backbone and 78.2% on the biomedical backbone—roughly a sixteen-fold higher refusal rate on unfamiliar-center data (Table VIII). The system therefore defers the overwhelming majority of out-of-center inputs to a clinician rather than asserting predictions on them, realizing the third of the three properties. Treated as a standard out-of-distribution detection task, the Mahalanobis score separates the unseen center sharply: AUROC 0.968 and FPR at 95% recall 0.185 on the dental backbone, and AUROC 0.951 with FPR95 0.306 on the biomedical backbone. The comparison uses matched tooth crops on both sides—Tufts crops extracted from its segmentation annotations—so the separation reflects genuine acquisition-center shift rather than a difference in how the two centers are sampled; computing the same comparison with a backbone-independent grid sampling of the Tufts dentition band yields the same refusal rates to within a fraction of a percent, confirming the effect is not an artifact of the sampling procedure. Against logit-based detectors on the same comparison, the embedding-space gate is far stronger (Table VII): on the dental backbone it reaches AUROC 0.968 versus 0.78 for energy and 0.78 for maximum softmax probability and ODIN, and on the biomedical backbone 0.951 versus 0.55 for all three. The logit-based scores perform near-identically and weakly because the classifier is saturated on out-of-center inputs—it reports high confidence on Tufts teeth as readily as on DENTEX teeth, leaving the logits with little to distinguish—whereas the Mahalanobis score reads the embedding geometry directly and so abstains before any prediction is formed. Extending the analysis to a third center sharpens this picture. On ADCD, an external-validation set drawn from other clinics with acquisition closer to DENTEX than Tufts, the gate refuses 33% of inputs at the same source operating point that refuses 5% of DENTEX, against 85% on the more distant Tufts center (Table IX). The refusal rate thus increases monotonically with the magnitude of the center shift—DENTEX 5%, ADCD 33%, Tufts 85%—which is precisely the behavior expected of a calibrated out-of-distribution gate: it abstains more as inputs move further from the training distribution, rather than refusing arbitrarily. On both external centers, per-center threshold recalibration using only a few dozen unlabeled target images reduces refusal to a usable level without retraining; on ADCD, accepted-case reliability could not be evaluated because its annotations contain only pathological findings and no labeled normal teeth to serve as negatives, so a definitive accepted-case evaluation requires a prospective per-tooth-labeled cohort.

TABLE VII

OUT-OF-DISTRIBUTION DETECTION ON TUFTS (MATCHED TOOTH CROPS), DENTEX-TEST VS TUFTS. THE EMBEDDING-BASED MAHALANOBIS GATE SUBSTANTIALLY OUTPERFORMS LOGIT-BASED DETECTORS (MSP, ENERGY, ODIN), WHICH COLLAPSE TOGETHER BECAUSE THE CLASSIFIER IS OVERCONFIDENT ON OUT-OF-CENTER INPUTS.

Detector	Dental AUROC	Dental FPR95	Biomed AUROC	Biomed FPR95
MSP	0.775	0.650	0.553	0.926
Energy	0.786	0.650	0.553	0.926
ODIN	0.775	0.650	0.553	0.926
Mahalanobis (ours)	0.968	0.185	0.951	0.306

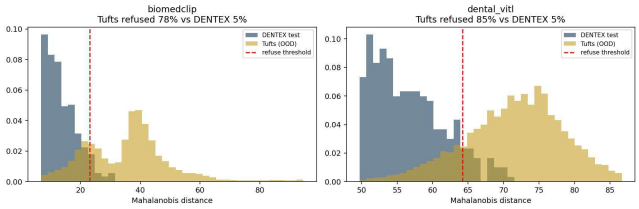


Fig. 5. Cross-center out-of-distribution refusal. At a threshold refusing 5% of DENTEX (blue), the envelope refuses 78% (BiomedCLIP) and 85% (dental ViT-L) of the unseen Tufts center (gold), using matched tooth crops on both sides.

TABLE VIII

CROSS-CENTER REFUSAL ON TUFTS USING MATCHED TOOTH CROPS. AT A 5%-ON-DENTEX OPERATING POINT, THE ENVELOPE REFUSES THE LARGE MAJORITY OF AN UNSEEN CENTER; MEAN MAHALANOBIS DISTANCE SHIFTS UPWARD ON BOTH BACKBONES.

Backbone	DENTEX refused	Tufts refused	Dist. shift (mean)
Dental ViT-L	5.1%	84.7%	51.9 → 71.6
BiomedCLIP	5.1%	78.2%	11.7 → 36.2

TABLE IX

REFUSAL RATE ACROSS THREE CENTERS AT THE SOURCE (DENTEX 95TH-PERCENTILE) OPERATING POINT. THE GATE'S REFUSAL RATE INCREASES MONOTONICALLY WITH THE MAGNITUDE OF THE CENTER SHIFT.

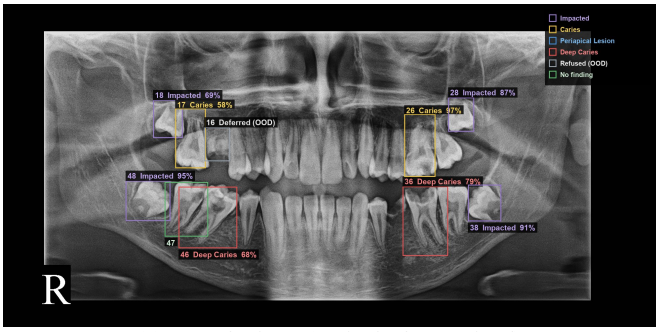
Center	Shift vs. DENTEX	Refusal rate
DENTEX (source)	— (in-distribution)	5%
ADCD (other clinics)	mild	33%
Tufts (other country)	severe	85%

H. Clinical translation: from calibrated findings to a safety-preserving report

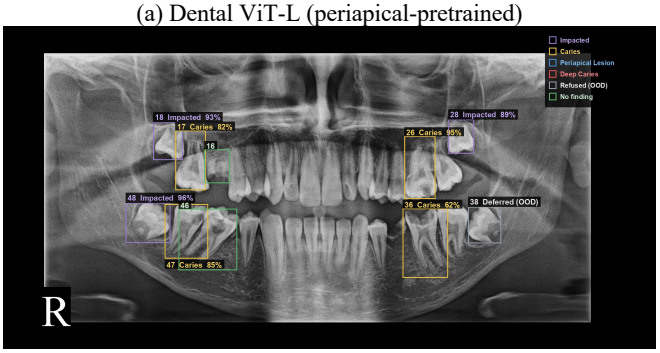
The envelope is intended to drive a clinician-facing report, and the three guarantees are precisely what make such a report defensible. We illustrate the deployment surface on a representative DENTEX case under both backbones (Fig. 6). For each tooth the calibrated envelope produces a probability per condition; a finding is stated only when the tooth is *accepted* by the out-of-distribution gate and its calibrated probability exceeds a reporting threshold. Accepted findings are rendered on the panoramic with their condition and calibrated confidence, while teeth the gate refuses are marked as deferred and are never assigned a finding. The same accept/refuse decision evaluated statistically in Section V-G thus becomes, at the level of an individual patient, the difference between a stated finding and an explicit deferral to the clinician.

The same case is shown under both backbones. Each accepts most teeth and reports a range of calibrated confidences, from high-confidence impactions to near-threshold caries calls. The two encoders agree on the difficult calls in both directions: both flag tooth 17 as caries—at 58% and 82% confidence—on what our clinical co-author read as a likely artifact, and both leave tooth 35, a root remnant, unmarked because it falls outside the four-condition label space. They differ elsewhere: the periapical backbone flags tooth 46 as deep caries and defers tooth 16, whereas the panoramic-capable backbone marks both as sound, missing a carious tooth and a root remnant. Neither backbone is uniformly better, and the errors that remain—a

shared false positive, a shared out-of-scope omission, and a handful of calls that move between the two—sit at the level of the task and the label set rather than the encoder, consistent with the aggregate result in Section VII. The graded confidences make the calibration property visible: the report asserts only what the bounded predictions support, surfaces every refusal, and stops at findings, listing the clinical confirmations needed before any treatment decision rather than recommending treatment. We offer it as a demonstrator of the deployment surface, not a validated reporting study; a reader study with practicing dentists across several centers is the natural next step and is left to future work.



(a) Dental ViT-L (periapical-pretrained)

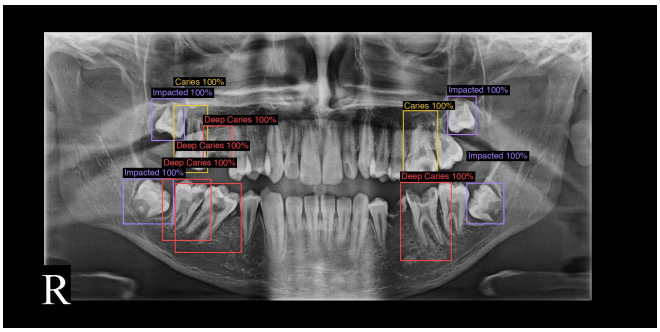


(b) DINOv3-L (panoramic-capable)

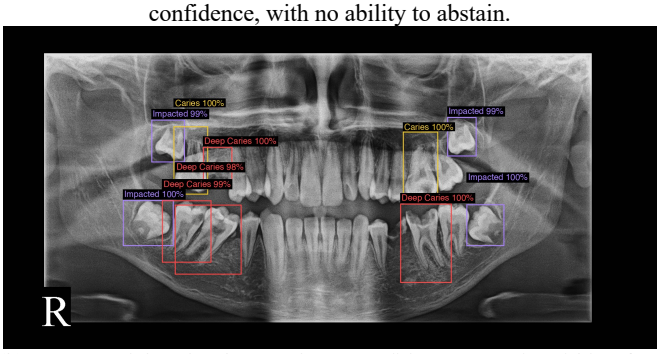
Fig. 6. Safety-enveloped findings on train_564 under two backbones: (a) Dental ViT-L and (b) DINOv3-L, each with its own calibrated envelope. Accepted teeth carry FDI number, condition, and calibrated confidence; teeth above the Mahalanobis threshold are deferred. Both backbones behave similarly (including a shared tooth-17 false positive); see Section V-H. Reviewed by the clinical co-author.

I. Behavior on a task-trained backbone

The envelope is conservative by construction, and a task-trained backbone makes this explicit. We attached the same envelope to a high-accuracy convolutional model (EfficientNet-V2-S, the strongest of the Table I zoo) by treating its penultimate features as a frozen embedding. Because that model was trained end-to-end on these conditions, its calibrated baseline is already sharply peaked: the temperature fit returns T = 3.4, and almost every finding is asserted at or near 100% confidence (Fig. 7a). Here the envelope has little room to act. The bounded residual keeps the final prediction within delta of this confident baseline, so the reported confidences move only marginally (100% to 98-100%, Fig. 7b), no tooth is deferred, and discrimination is unchanged (macro-AUC 0.909 to 0.909). Calibration is essentially unchanged and not improved (ECE 0.056 to 0.062). This is the intended behavior, not a shortfall: the bounded-deviation contract guarantees the envelope cannot drift far from a strong baseline, so on an already-confident model it stays faithful rather than perturbing it. The calibration benefit reported earlier instead appears where the embedding geometry leaves room for it, on the diffuse frozen foundation embeddings (dental ViT-L, ECE 0.040 to 0.034). The practical reading is that the envelope supplies bounded robustness on any backbone while delivering calibration gains where the feature geometry permits and never degrades a strong discriminator.



(a) Raw EfficientNet-V2-S: every finding asserted at 100% confidence, with no ability to abstain.



(b) Same model under the envelope: confidences remain within ±δ of the confident baseline (100% to 98-100%); the bound keeps the strong baseline intact rather than recalibrating it.

Fig. 7. The envelope on a task-trained backbone (EfficientNet-V2-S), train_564. (a) The raw model asserts every finding at 100% with no deferral; (b) under the envelope, confidences stay within ±δ of this already-confident baseline (AUC 0.909→0.909), as discussed in Section V-I. Confidences are uncalibrated model outputs, not validated diagnoses.

VI. Discussion

Three findings deserve emphasis. First, the value of the envelope is *conditional on the backbone*: it reduces calibration error substantially on the dental encoder that needs it and harmlessly preserves the already-calibrated biomedical encoder. “Improves where needed, never harms, with a formal guarantee” is a stronger and more honest claim than a uniform accuracy gain, and is the claim the data support. Second, the safety guarantee is *architectural*: it holds in all eighteen tested conditions because it is a property of the clip, not of the data, and is therefore not subject to the usual concerns about out-of-sample generalization. Third, the refusal mechanism is *agnostic in form but backbone-dependent in strength*—it works on both encoders but separates the unseen center far more sharply on the dental one, indicating that better embedding geometry

yields better refusal. The same agnosticism shows up in the backbone itself: replacing the periapical encoder with a modern panoramic-pretrained generalist left the in-distribution error rate essentially unchanged under a matched probe, which points back to the safety layer, not the encoder, as the part that does the work. As stronger dental foundation models are released, the same envelope is expected to inherit their improved geometry without modification, which is the central practical advantage of a frozen-backbone, drop-in design. This mirrors the trajectory of foundation models in other imaging specialties, where domain-specific encoders such as RETFound [23] and the broader generalist-medical-AI program [41], [24] have rapidly displaced task-specific networks; a portable safety layer ensures each such advance can be deployed responsibly.

VII. Limitations

The discrimination gains from the envelope are modest by design; the contribution is safety, not accuracy, and the paper should not be read as a performance result. The failure modes the envelope targets—silent degradation, miscalibration, and unseen-center inputs—are precisely those documented as clinically consequential in the medical-imaging literature [21], [42], [43]. DENTEX is small (705 images), and the rarest condition, Periapical Lesion (158 instances), remains the noisiest. The cross-center evaluation uses two external centers (Tufts and ADCD); both are public datasets, and a prospective clinical cohort with per-tooth normal/abnormal labels would further strengthen the generalization claim and enable a definitive accepted-case reliability evaluation, which is the most valuable next step. The dental encoder was trained on periapical images and applied to panoramic crops, a domain shift we treat openly rather than hide. We checked whether this mismatch is what drives the errors: running a panoramic-capable generalist backbone (DINOv3-L) on the same matched teeth, under an identical linear-probe protocol, gave discrimination indistinguishable from the periapical encoder. The residual errors are therefore task-level, not a by-product of the periapical-to-panoramic shift. Finally, the bounded-deviation guarantee is stated relative to a simple calibrated baseline; substituting a clinician-derived prior would make the contract clinically, not merely statistically, meaningful. The annotated report consumes tooth bounding boxes from the benchmark annotations; the envelope classifies tooth crops and does not localize teeth, so deployment on a new panoramic without annotations requires a separate tooth-detection stage upstream. We do not benchmark against ensemble-based uncertainty methods such as deep ensembles or Monte-Carlo dropout: these require either multiple independently trained networks or stochastic forward passes through a trainable backbone, which is incompatible with the frozen, single-encoder, no-retraining setting that defines the method and motivates its low cost; the post-hoc detectors that are compatible with a frozen encoder (temperature scaling, and Mahalanobis-based OOD scoring) are the ones we adopt, and a broader head-to-head against compatible post-hoc detectors such as ODIN and energy scores is a natural extension. None of these limitations affects the structural guarantee, which holds regardless of data scale or center.

VIII. Conclusion

We presented a backbone-agnostic safety envelope that equips frozen dental foundation embeddings with calibrated confidence, a provable bounded-deviation contract that recovers a trusted baseline in the worst case, and out-of-distribution refusal. After showing that accurate task-trained dental classifiers are routinely miscalibrated, we demonstrated, across two structurally different encoders and three replicated seeds, that the envelope improves calibration where the backbone needs it while preserving discrimination and holding its guarantee in every condition; on a genuinely different clinical center it defers 85% of inputs at an operating point that defers only 5% of its own. The layer is small, portable, and immediately reusable as stronger dental foundation models become available, adding calibrated confidence, a bounded-deviation contract, and refusal to an otherwise opaque predictor.

Acknowledgment

The authors thank the organizers of the DENTEX challenge and the creators of the Tufts Dental Database for making their panoramic radiograph datasets publicly available for research; this work would not have been possible without these resources. The authors also acknowledge the developers of the open-source foundation models and software used in this study, including the dental visual-expert and BiomedCLIP encoders, DINOv3, and the PyTorch, Hugging Face, and scikit-learn libraries.

During the preparation of this work, the author used generative AI tools (Claude, ChatGPT, and Gemini) to assist with reviewing and editing the manuscript and with figure drafting. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the accuracy and integrity of the final publication.

References

[1] C. Du, P. Li, X. Wang, and Y. Zhang, “Prompting vision-language models for dental notation aware abnormality detection,” in Proc. MICCAI, 2024, pp. 687–697.

[2] J. Hao, X. Wang, B. Liu, and Y. Zhang, “OralGPT-Omni: A versatile dental multimodal large language model,” arXiv:2511.22055, 2025.

[3] X. Huang, J. Li, Y. Chen, and L. Zhang, “Towards generalist intelligence in dentistry: Vision foundation models for oral and maxillofacial radiology,” arXiv:2510.14532, 2025.

[4] Y. Geifman and R. El-Yaniv, “Selective classification for deep neural networks,” in Proc. NeurIPS, vol. 30, 2017.

[5] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, “Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm,” J. Dentistry, vol. 77, pp. 106–111, 2018.

[6] J. Krois, T. Ekert, L. Meinhold, T. Golla, B. Kharbot, A. Wittemeier, C. Dörfer, and F. Schwendicke, “Deep learning for the radiographic detection of periodontal bone loss,” Sci. Rep., vol. 9, 8495, 2019.

[7] F. Schwendicke, W. Samek, and J. Krois, “Artificial intelligence in dentistry: Chances and challenges,” J. Dental Res., vol. 99, no. 7, pp. 769–774, 2020.

[8] T. H. Bui, K. Hamamoto, and M. P. Paing, “Automated caries screening using ensemble deep learning on panoramic radiographs,” Entropy, vol. 24, no. 10, 1358, 2022.

[9] B. Dayı, H. Üzen, İ. Balıkçı Çiçek, and Ş. B. Duman, “A novel deep learning-based approach for segmentation of different type caries lesions on panoramic radiographs,” Diagnostics, vol. 13, no. 2, 202, 2023.

[10] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, “A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images,” IEEE Trans. Med. Imag., vol. 39, no. 8, pp. 2653–2663, 2020.

[11] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, “A smart dental health-IoT platform based on intelligent hardware, deep learning,

and mobile terminal," IEEE J. Biomed. Health Inform., vol. 24, no. 3, pp. 898–906, 2020.
[12] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. ICML, 2017, pp. 1321–1330.
[13] M. Minderer, J. Djolonga, R. Romijnders, F. Hubis, X. Zhai, N. Houlsby, D. Tran, and M. Lucic, "Revisiting the calibration of modern neural networks," in Proc. NeurIPS, vol. 34, pp. 15682–15694, 2021.
[14] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in Proc. ICLR, 2017.
[15] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in Proc. NeurIPS, vol. 31, 2018.
[16] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Proc. NeurIPS, vol. 30, 2017.
[17] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. ICML, 2016, pp. 1050–1059.
[18] Y. Ovadia et al., "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in Proc. NeurIPS, vol. 32, 2019.
[19] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in Proc. ICLR, 2018.
[20] C. Tomani, S. Gruber, M. E. Erdem, D. Cremers, and F. Buettner, "Post-hoc uncertainty calibration for domain drift scenarios," in Proc. IEEE/CVF CVPR, 2021, pp. 10124–10132.
[21] C. Zhou, S. Wang, and Q. Liu, "Out-of-distribution detection in medical image analysis: A survey," arXiv:2404.18279, 2024.
[22] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. ICML, 2021, pp. 8748–8763.
[23] Y. Zhou et al., "A foundation model for generalizable disease detection from retinal images," Nature, vol. 622, no. 7981, pp. 156–163, 2023.
[24] M. Awais et al., "Foundation models defining a new era in vision: A survey and outlook," IEEE Trans. Pattern Anal. Mach. Intell., vol. 47, no. 4, pp. 2245–2264, 2025.
[25] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.
[26] A. Elnashar, "Dual-space trust-region active-set optimization for robust LCMV beamforming with multiple worst-case constraints," Research Square preprint, doi:10.21203/rs.3.rs-9807253/v1, 2025.
[27] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World. New York, NY, USA: Springer, 2005.
[28] A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," arXiv:2107.07511, 2021.
[29] A. N. Angelopoulos, S. Bates, A. Fisch, L. Lei, and T. Schuster, "Conformal risk control," arXiv:2208.02814, 2022.
[30] S. Bates, A. Angelopoulos, L. Lei, J. Malik, and M. Jordan, "Distribution-free, risk-controlling prediction sets," J. ACM, vol. 68, no. 6, pp. 1–34, 2021.
[31] I. E. Hamamci, S. Er, E. Simsar, A. Sekuboyina, M. Gundogar, B. Stadlinger, A. Mehl, and B. Menze, "DENTEX: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic X-rays," arXiv:2305.19112, 2023.
[32] K. Panetta, R. Rajendran, A. Ramesh, S. P. Rao, and S. Agaian, "Tufts Dental Database: A multimodal panoramic X-ray dataset for benchmarking diagnostic systems," IEEE J. Biomed. Health Inform., vol. 26, no. 4, pp. 1650–1659, Apr. 2022.
[33] S. Zhang et al., "BiomedCLIP: A multimodal biomedical foundation model pretrained from fifteen million image–text pairs," arXiv:2303.00915, 2023.
[34] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," in Proc. ICML, 2021, pp. 10096–10106.
[35] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 4700–4708.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
[37] A. Mehrtash, W. M. Wells, C. M. Tempany, P. Abolmaesumi, and T. Kapur, "Confidence calibration and predictive uncertainty estimation for deep medical image segmentation," IEEE Trans. Med. Imag., vol. 39, no. 12, pp. 3868–3878, 2020.
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.
[39] F. Schwendicke et al., "Convolutional neural networks for dental image diagnostics: A scoping review," J. Dentistry, vol. 91, 103226, 2019.
[40] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," Med. Image Anal., vol. 79, 102444, 2022.
[41] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," Nature, vol. 616, no. 7956, pp. 259–265, 2023.
[42] L. Oakden-Rayner, J. Dunnmon, G. Carneiro, and C. Ré, "Hidden stratification causes clinically meaningful failures in machine learning for medical imaging," in Proc. ACM Conf. Health, Inference, Learning, 2020, pp. 151–159.
[43] M. Roberts et al., "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," Nat. Mach. Intell., vol. 3, pp. 199–217, 2021.
[44] Annotated Dental Conditions Dataset (ADCD): panoramic radiography YOLO dataset, 14 classes, Zenodo, record 15487430, 2024. [Online]. Available: https://doi.org/10.5281/zenodo.15487430



AYMAN ELNASHAR received his B.Sc., M.Sc., and Ph.D. degrees in Electrical Communications Engineering (1995, 1999, 2005) and holds an Executive MBA. He has more than 30 years of combined industry and academic experience spanning senior leadership roles, applied research, and teaching. He is currently Vice President - Technology Strategy, Architecture & Innovation at e& UAE (etisalat), and Adjunct Professor at Khalifa University. In parallel, he serves as an Advisory Board Member at Lablabee, driving innovation and strategic oversight in technology education.

Dr. Elnashar previously held senior leadership roles at du (UAE), Mobily (KSA), and Orange (Egypt), where he contributed to the design and rollout of large-scale mobile, fixed, and cloud infrastructures. He also plays a prominent role in the academic and innovation ecosystem as an Industry Advisory Board Member at NYU Abu Dhabi, RIT Dubai, Canadian University Dubai and AUS. He has authored three Wiley reference books and an extensive body of peer-reviewed publications and industry whitepapers, spanning applied 6G/5G/4G/IoT system design and theoretical advances in adaptive detection and beamforming, digital signal processing, and neural networks. His research interests span wireless networks, AI/ML, AI-native networks, cloud-native systems, smart cities, and IoT, with a focus on translating advanced theory into practical, real-world ICT solutions.



Norsin Elnashar received the Bachelor of Dental Surgery (BDS) degree with honors from Ajman University, Ajman, United Arab Emirates, in 2024, where she also completed a 12-month accredited dental internship across all eight core clinical disciplines. She subsequently received the Diploma in Primary Care Dentistry (Dip PCD) from the Royal College of Surgeons in Ireland (RCSI), Dublin, Ireland, and completed the Patient Safety and Quality Leadership Specialization from Johns Hopkins University, Baltimore, MD, USA. She is currently a Clinical Instructor in Oral Surgery with the College of Dentistry, Ajman University. Her research interests include evidence-based clinical dentistry, patient safety, and the application of data-driven and imaging-based methods to dental clinical practice.



Ahmed Al-Emam received the M.B.B.Ch. degree in 1998 and the M.Sc. degree in forensic medicine and clinical toxicology in 2004, both from Mansoura University, Egypt, and the Ph.D. degree in medical and molecular genetics from the University of Birmingham, U.K., in 2011, where his research addressed DNA repair and genome stability.

He is currently a Professor of Forensic Medicine and Clinical Toxicology at Mansoura University, Egypt. Since November 2012, he has been with King Khalid University, KSA, where he serves as Course Coordinator of Forensic Medicine and Clinical Toxicology and contributes to teaching Applied Molecular Medicine. He is a member of the committee for course specification and development of Applied Molecular Medicine. His research interests include DNA repair, genome stability, medical and molecular genetics, and clinical toxicology. He has authored a substantial body of peer-reviewed publications in these areas.