UNIVERSIDAD EARTH

Can a *Phone-Sized Model* Replace a Research-Grade One?

A pre-registered equivalence test of **SegFormer (47 M)** *against* **FastViT-T8 (3.6 M)** *with DINOv2 distillation for SENASA coffee-disease grading*

COFFEA ARABICA VAR. ESPERANZA L4A5 · CARIBBEAN LOWLANDS OF COSTA RICA

Carly Chery

under the direction of **Victor Hugo Morales Peña, Ph.D.**

EARTH UNIVERSITY · GUÁCIMO, LIMÓN · COSTA RICA

PROTOCOL ON OSF · OSF.IO/E7N6J · DEFENSE DECEMBER 2026

3.6 M	10×	±0.05	1,400
FastViT-T8 · *phone-sized*	*smaller than the segmenter*	*equivalence margin (κ)*	*annotated coffee leaves*

CONTRIBUTIONS

Principal claims

i. **Preregistered equivalence, not superiority.** *SegFormer* (47 M) vs. *FastViT-T8 + DINOv2* distillation (3.6 M) for SENASA coffee-disease grading. Three baselines (ResNet-18, U-Net, DINOv2 linear probe) and four hypotheses locked on OSF before data collection.

ii. **TOST declares "equal", not "no difference".** Two one-sided tests (Schuirmann, 1987; Lakens, 2017) let us affirmatively claim a 10× smaller classifier the practical equal of a research-grade segmenter to within ±0.05 κ — a finding classical superiority testing cannot deliver.

iii. **Field-ready pipeline.** Offline *Gradio* laptop app (Spanish UI) for farm technicians now; *LiteRT* (Android) and *Core ML* (iOS) mobile exports prepared for Phase 2.

Coffee leaf rust (*Hemileia vastatrix*) can cost up to fifty percent of a crop's yield; American leaf spot (*Mycena citricolor*) thrives in Caribbean-lowland humidity. Severity is still graded by eye on the subjective SENASA 0–3 scale. This study preregisters a head-to-head comparison of **SegFormer** (MiT-B3, 47 M params; pixel-wise segmentation) against **FastViT-T8** (3.6 M params; ordinal classification via DINOv2 LoRA distillation + CORN loss) on *Coffea arabica* var. Esperanza L4A5.

IMPACT

US $1–2 B / *year*

Estimated annual losses across Latin America attributable to H. vastatrix (Avelino et al., 2015). A laptop-deployable grader enables early, low-cost intervention for smallholder farmers.

NOVELTY

A pre-specified comparison of SegFormer against FastViT T8 with DINOv2 LoRA distillation on the SENASA coffee-severity scale, framed under a TOST equivalence hypothesis against ResNet-18, U-Net, and DINOv2 linear-probe baselines.

III. Objective

the deliverable we commit to building

To deliver a precise, consistent, and reproducible severity grader for rust and leaf-spot diseases on Coffea arabica var. Esperanza L4A5 (first as an offline Gradio laptop application for field technicians, subsequently exported to mobile inference), and to determine, by preregistered statistical test, whether a four-million-parameter classifier can stand as the practical equivalent of a forty-five-million-parameter segmentation model.

IV. Methodology

five-stage pipeline from field capture to deployment

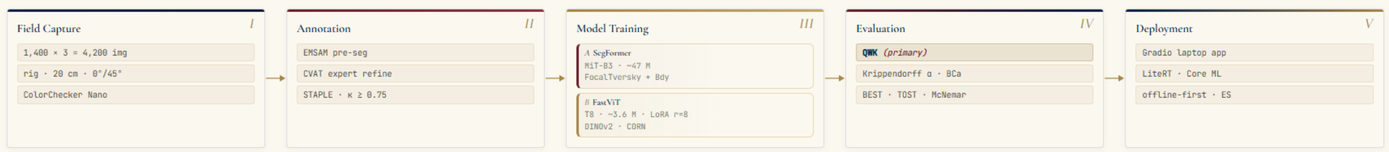


FIGURE I. *End-to-end methodological pipeline. Five stages from field capture on Coffea arabica var. Esperanza L4A5 (I), through semi-automated annotation with STAPLE consensus (II), parallel training of SegFormer and FastViT under preregistered hyperparameters (III), the preregistered comparative statistical framework (IV), to laptop-first Gradio deployment with mobile export readied (V).*

Plantation — *What is at stake economically — up to 50% yield loss at epidemic peak.*

Leaf (≈ grade G2) — *What the model classifies on the SENASA 0–3 scale.*

Pustule — *What the pixel-wise segmentation learns to recognise.*

FIGURE II. *The same problem at three scales — plantation, leaf, and pustule — each the unit of a different stakeholder decision. Photos: Wikimedia Commons, CC BY-SA 3.0; not from this study's dataset.*

V. Dataset

1,400 × 3

unique leaves × orientations = ***4,200 images*** *· only the 1,400 adaxial are annotated*

Healthy (shared)	200
H. vastatrix G1 · G2 · G3	3 × 200
M. citricolor G1 · G2 · G3	3 × 200
Total · 7 categories	1,400

Power: Monte Carlo n = 1,400 at ≥ 0.90 for Cohen's w = 0.3 on Stuart-Maxwell. Leaf-level grouping precludes CV leakage; 2,800 oblique + abaxial reserved for augmentation & robustness. Hybrid scoring: ordinal metrics (QWK, α) computed within each disease; disease-type recognition (rust / spot / healthy) scored as separate multiclass accuracy.

VI. Preregistered Targets

METRIC	METHOD	TARGET
▸ ***Quadratic Weighted κ*** PRIMARY	both	≥ 0.80
Mean IoU (Jaccard)	SegFormer	≥ 0.75
Dice coefficient	SegFormer	≥ 0.80
Boundary IoU 2PX/10 PX	SegFormer	report
Per-class F_1	both	≥ 0.85
Krippendorff's α ORDINAL · QUADRATIC · VS EXPERTS	both	≥ 0.80
ECE (15 bins) CALIBRATION	both	report
Latency P50 / P95 / P99	both	MLPerf

Baselines reported alongside hero methods: ResNet-18 + CE · U-Net + compound loss · DINOv2 linear probe.

VII. Preregistered Hypotheses

H_1 Superiority *(Stuart-Maxwell · two-sided $\alpha = 0.05$)*

$\kappa_{SegFormer} \neq \kappa_{FastViT}$

H_2 Non-inferiority *(one-sided $\alpha = 0.025$ · $\delta = 0.05$)*

$\kappa_{FastViT} > \kappa_{SegFormer} - 0.05$

● PRIMARY TEST

H_3 TOST equivalence *(two one-sided $\alpha = 0.025$ each · family $\alpha = 0.05$)*

$|\kappa_{SegFormer} - \kappa_{FastViT}| \leq 0.05$

H_4 Efficiency *(conditional on H_2 or H_3)*

$FLOPs_{FastViT} \leq FLOPs_{SegFormer} / 10$

$Memory_{FastViT} \leq Memory_{SegFormer} / 10$

∂ = 0.05 κ is below typical inter-rater disagreement on SENASA coffee grading (κ ≈ 0.60–0.75; Esgario et al., 2020; Tassis et al., 2021). Family-wise error controlled by **Holm-Bonferroni** over H1–H3; H4 conditional. BEST (Kruschke, 2013) with weakly-informative priors; BCa bootstrap, 10,000 resamples.

VIII · EXPERT VALIDATION PROTOCOL

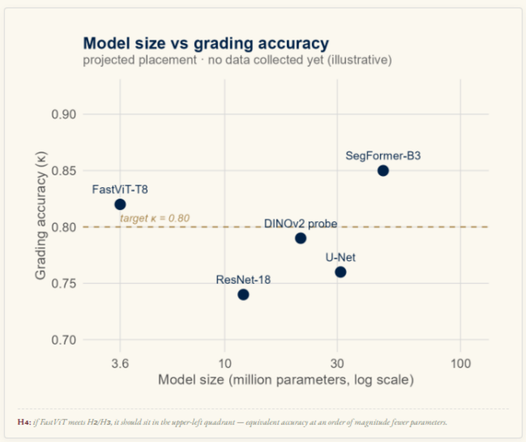
Three phytopathologists · blind grading · 210 stratified images

Independent classification across the seven severity categories, with agreement assessed by Krippendorff's α (ordinal) and Bland–Altman limits. Inter-expert α defines a practical ceiling reflecting the irreducible annotator disagreement characteristic of the SENASA scale.

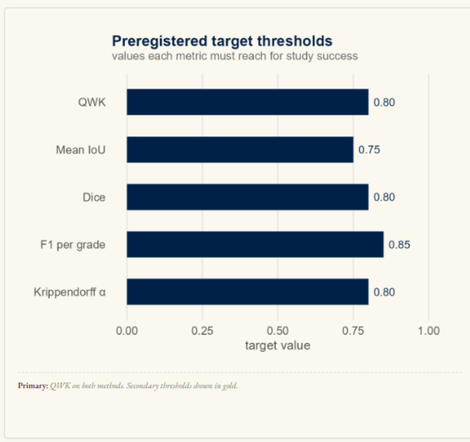
3	210	$\alpha_{\geq 0.80}$
EXPERT PHYTOPATHOLOGISTS	IMAGES · 30 PER CATEGORY	KRIPPENDORFF (ORDINAL) TARGET

IX. Expected Results

preregistered targets visualised

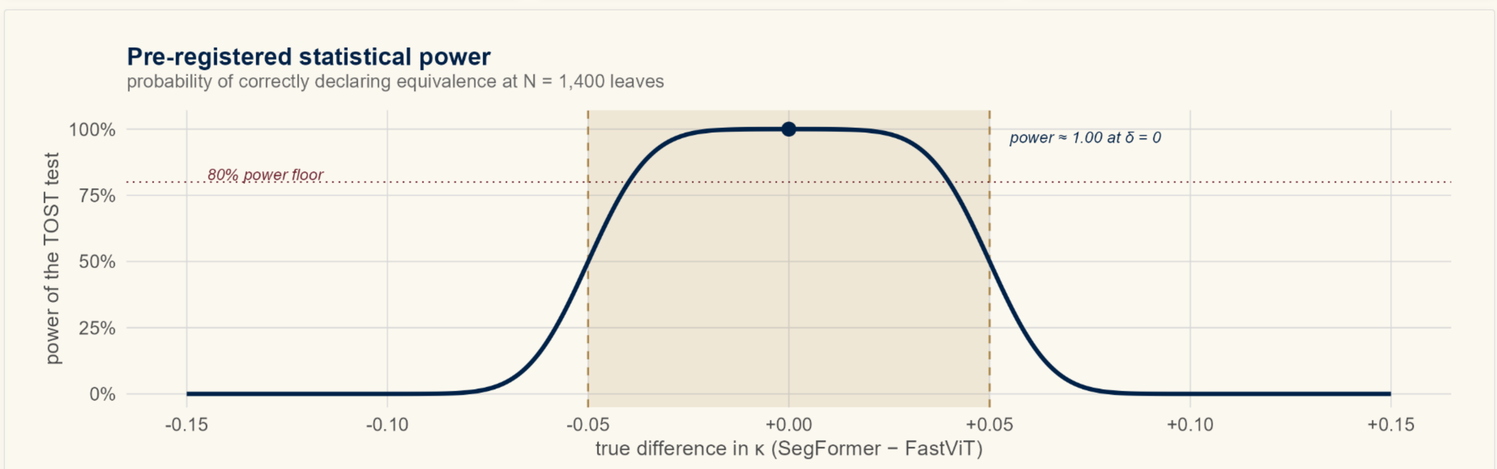


H4: *if FastViT meets H2/H3, it should sit in the upper-left quadrant — equivalent accuracy at an order of magnitude fewer parameters.*



Primary: *QWK on both methods. Secondary thresholds shown in gold.*

H2: *95 % CI of $\kappa_{Seg} - \kappa_{Fast}$ must lie entirely within ±0.05 to declare equivalence.*



Design choice: *N = 1,400 leaves was chosen so the TOST test has > 80% power to correctly declare equivalence when the true difference in κ is within ±0.03 — well below the ±0.05 margin. The sample size is not a convenience; it is a design parameter, fixed before any data collection.*

X. Recommendations

three verbs for the next phase

X. i

Deploy.

FastViT-T8 as an *offline Gradio laptop app* (Spanish UI). Mobile exports — *LiteRT* (Android) and *Core ML* (iOS) — readied for Phase 2.

X. ii

Generalise.

Extend beyond *Esperanza L4A5* via farmer-cooperative partnerships. SegFormer handles co-infection natively once training data on dual-infected leaves exists.

X. iii

Trust.

Publish segmentation overlays with every grade so an agronomist can audit *why*. Governance must ensure the tool *serves* farmers, never surveils them.

XI. Limitations & Scope

Single site (EARTH Forestal Farm) · **single cultivar** (*Coffea arabica* var. Esperanza L4A5) · **co-infected leaves out of scope** · **single season** (inter-annual drift not evaluated) · **field-usability study deferred** · **pilot empirical results pending** (Q2–Q3 2026).

SELECTED REFERENCES

1. Avelino, J. *et al.* (2015). The coffee rust crises in Colombia and Central America (2008–2013). *Food Security, 7*, 303–321.
2. Esgario, J.G.M. *et al.* (2020). Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric., 169*, 105162.
3. Kruschke, J.K. (2013). Bayesian estimation supersedes the t-test (BEST). *J. Exp. Psychol. Gen.*, 142, 573–603.
4. Lakens, D. (2017). Equivalence tests: a practical primer. *Soc. Psychol. Pers. Sci.*, 8, 355–362.
5. Li, J. *et al.* (2025). EMSAM: enhanced multi-scale segment anything model for leaf disease segmentation. *Frontiers in Plant Science*.
6. Oquab, M. *et al.* (2024). DINOv2: learning robust visual features without supervision. *TMLR*.
7. Schuirmann, D.J. (1987). A comparison of the two one-sided tests procedure for equivalence of bioavailability. *J. Pharmacokinet. Biopharm.*, 15, 657–680.
8. Shi, X. *et al.* (2021). Deep neural networks for rank-consistent ordinal regression (CORN). *arXiv:2111.08851*.
9. Tassis, L.M. *et al.* (2021). Instance + semantic segmentation of coffee-leaf diseases and pests. *Comput. Electron. Agric.*, 186, 106191.
10. Vasu, P.K.A. *et al.* (2023). FastViT: a fast hybrid vision transformer using structural reparameterization. *ICCV*.
11. Warfield, S.K., Zou, K.H. & Wells, W.M. (2004). Simultaneous truth and performance level estimation (STAPLE). *IEEE TMI*, 23, 903–921.
12. Xie, E. *et al.* (2021). SegFormer: simple and efficient design for semantic segmentation with transformers. *NeurIPS*.

PROJECT · PROTOCOL · FILES

osf.io/e7n6j

In memoriam Maman · *with thanks to the Morales Peña laboratory and the EARTH Forestal Farm team*

DOCKER · DVC · W&B · OPTUNA · OSF · MODEL CARDS · DATASHEETS · DATA CC BY 4.0 · CODE MIT · WEIGHTS APACHE 2.0 · COMPUTE DISCLOSURE AT RELEASE

#DeepLearning · #CoffeeDiseaseAI · #PlantPhenotyping · #EARTHUniversity · #PG26038