Hierarchical Equiangular Tight Frames for Matching Relief-Printed Patterns on Medieval Ceramic Sherds

Yassine Nasser[1][0000−0002−5347−0386], Sylvie Treuillet[1], Matthieu Exbrayat[2], and Sébastien Jesset[3]

[1] PRISME Laboratory, UR 4229, University of Orléans, France
{yassine.nasser,sylvie.treuillet}@univ-orleans.fr
[2] LIFO, UR 4022, University of Orléans, France
matthieu.exbrayat@univ-orleans.fr
[3] Service Archéologique Municipal d'Orléans, France
Sebastien.Jesset@orleans-metropole.fr

Abstract. Matching relief-printed decorations on medieval ceramic sherds to the wooden thumbwheels that produced them is essential for reconstructing ancient pottery workshops and trade networks. The task is intrinsically difficult: only a few sherds can be authoritatively associated with a given wheel by archaeologists, and the corpus exhibits severe class imbalance, erosion-induced intra-class variability, and strong inter-class similarity within typological families. The decoration corpus follows a two-level hierarchy: each thumbwheel is assigned to a coarse typological super-class defined by the potter's engraving gesture (Squares G, Squares H, Chevrons L, Diamonds A), within which fine-grained sub-classes capture individual wheel identities. We propose a hierarchical recognition framework in which the classifier geometry is fixed a priori as a *Hierarchical Equiangular Tight Frame* (H-ETF). A global ETF models the four super-classes, while local ETFs encode wheel identities within mutually orthogonal subspaces. The construction decomposes the embedding so that Neural Collapse, when attained by the trained features, holds at both hierarchy levels simultaneously. Coupled with a self-supervised DINOv2 ViT-S/14 backbone with partial fine-tuning, recognition reduces to nearest-vertex assignment in the H-ETF embedding space. On the REMIA project corpus (1,123 images, 11 thumbwheels, 4 super-classes), H-ETF achieves $96.9 \pm 0.4\%$ sub-class accuracy and $99.9 \pm 0.1\%$ super-class accuracy under combined-source supervision, surpassing a same-backbone softmax baseline by +2.1pp and a strong EfficientNet-B5 baseline by +3.6pp. Under reduced-support evaluation on real shards (4-way, $K \in \{1, 3, 5\}$), H-ETF outperforms all baselines at every K, with the largest gap at K=1 (+6.6pp). Qualitative analyses confirm that the orthogonal subspace structure separates macro-pattern identity from fine micro-texture, in line with archaeological reasoning.

Keywords: Equiangular tight frame · Hierarchical classification · Archaeological ceramics · Engraved pottery sherds · Thumbwheel identification

Fig. 1: Left: excavated sherds from the Médecinerie site (Saran, Loiret) bearing relief-printed decorations. Right: an example of a bronze thumbwheel from late Antiquity, used to imprint geometric friezes in fresh clay.

1 Introduction

Discovered in 1968, the Médecinerie site at Saran (Loiret, France) is unique in France for its Merovingian and Carolingian (6th–9th century) layers. Excavations conducted by the Service Archéologique Municipal d'Orléans (SAVO) and the Fédération Archéologique du Loiret (FAL), as well as the Institut National de Recherches Archéologiques Préventives (INRAP) and the Service Départemental Archéologique du Loiret (SeDAL), revealed a major pottery production center, with numerous ceramic artefacts distributed across multiple kilns. We focus on the sherds bearing a relief frieze produced by the *thumbwheel* technique: a small cylinder—typically wood, sometimes bronze, iron, ceramic or bone—engraved with geometric patterns (diamonds, ovals, sticks) and rolled on the still-fresh clay (Fig. 1).

The material and natural wear of the thumbwheel determine the regularity of the printed motif. The pressure, speed and gesture used during application introduce further per-craftsman singularities. Combined with the rapid degradation of the wooden tool itself, each wheel records a set of micro-particularities that distinctively mark a single potter's production. These unique imprints make it possible to identify sherds produced by the same wheel, which is in turn central evidence for reconstructing ancient trade networks and the lifestyles of past civilizations.

The classical archaeological method to group sherds by wheel relies on *manual stamping*—applying modeling paste to the surface, inking, and visually superposing the resulting prints—a process that is both slow and tedious, leaving the corpus of associated sherds essentially unlabeled at scale. The aim of this work is to automate the recognition and grouping of sherds by their producing thumbwheel. Four difficulties make the problem genuinely hard:

(i) Severe class imbalance. Out of 1,123 images in the REMIA project corpus, the most populous sub-class (A2) contains 203 samples while the rarest (H12) contains only 20, a 10:1 ratio. Four sub-classes have fewer than 50 samples each.

(ii) Time-eroded sherds. Many real sherds are partially worn, with shallow or interrupted imprints, especially near fracture edges.

(iii) High intra-class variability. The wheel application itself is uneven: variations in rolling speed, slippage, clay texture, vase shape and curvature all alter the printed pattern from one sherd to the next, even when the same wheel is used.

(iv) High inter-class similarity. Conversely, distinct thumbwheels often share macro-geometric properties (e.g., all belong to the Squares family) and differ only by fine micro-texture: number of incisions, depth, degradation pattern. Brahim et al. [8] report that nearly all errors in their EfficientNet baseline concentrate within the Squares-H sub-family (H1, H12, H13).

The high intra-class variability and the high inter-class similarity together make a flat 11-way classifier inefficient: it must spend capacity separating wheels that already differ at the macro level (e.g., Squares versus Chevrons), at the cost of resolving the wheels that genuinely differ only at the micro level. Archaeologists themselves reason hierarchically: first recognize the engraving *gesture* (Squares-G, Squares-H, Chevrons, Diamonds), then look at fine micro-texture to attribute a specific wheel.

We mirror this hierarchical reasoning by fixing the classifier geometry, a priori, as a *Hierarchical Equiangular Tight Frame* (H-ETF). The Neural Collapse line of work [1, 2] has shown that ETFs are the optimal terminal geometry of cross-entropy classifiers and that pre-registering classifier weights as an ETF improves *imbalanced* and *low-data classification*. We extend this principle to two levels: a global ETF with 4 vertices encodes the super-classes, and a local ETF inside each super-class encodes its sub-classes, offset from the super-class centroid along an *orthogonal* subspace. The H-ETF head is mounted on a self-supervised DINOv2 ViT-S/14 backbone [7] with partial fine-tuning. At inference, classification reduces to nearest-vertex assignment.

We summarize our contributions as follows:

1. We introduce the H-ETF, a fixed two-level simplex geometry that encodes typological super-classes and thumbwheel sub-classes in mutually orthogonal subspaces, and show that this decomposition is compatible with Neural Collapse at both levels (Proposition 1).
2. We provide a benchmark on the REMIA project corpus over two complementary protocols: (P1) source-isolation classification (combined and real vs. facsimile), and (P2) low-data evaluation on real shards via reduced-support nearest-prototype classification. P1 measures the ceiling under each data regime; P2 isolates the low-data robustness that is critical when only a handful of real exemplars exist for a given wheel.

3. We show that H-ETF outperforms a same-backbone softmax baseline (+2.1pp on combined, +3.1pp on real) and exceeds a strong EfficientNet-B5 baseline (+3.6pp on combined, +6.4pp on real, despite using a smaller backbone). The advantage of H-ETF is most pronounced in the data-scarce real-only regime and at low support sizes.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the REMIA project corpus. Section 4 introduces the H-ETF construction and training. Section 5 presents the experimental protocols and results, and Section 6 discusses the findings and concludes.

2 Related Work

Computer-vision-assisted analysis of ceramics has evolved from hand-crafted shape and texture descriptors [9, 12–14] to deep learning. Early CNN-based work showed the value of learned features for pottery retrieval [20] and for classifying engraved medieval sherds by decoration *type* via compact bilinear pooling of CNN features [10, 11]. Subsequent efforts targeted Roman pottery, ranging from end-user identification tools [15–17] to unsupervised clustering of potsherds with a convolutional VAE [18]. Pawlowicz and Downum [19] applied VGG-style networks to fine-grained typology of Tusayan White Ware. Most relevantly, Brahim et al. [8] fine-tuned EfficientNet on 2D shaded views of 3D-scanned sherds, augmenting the training set with carved-wood facsimiles, and report up to 97.66% accuracy with an EfficientNet-B5 on the same REMIA project corpus we use.

These methods rely on (i) relatively large per-class samples—a condition that does not hold for archaeologically rare wheels—and (ii) a flat label structure that ignores the typological hierarchy archaeologists routinely use.

Papyan, Han and Donoho [1] showed that neural networks trained with cross-entropy enter a final regime called Neural Collapse: features from the same class cluster tightly around a single point, class representation centers arrange themselves as evenly spaced directions forming a simplex equiangular tight frame (ETF), and the classifier aligns with this geometric structure. Yang et al. [2] showed that an ETF classifier with a dot-regression loss is provably optimal for balanced classification *and* remains near-optimal under class imbalance and low-data conditions, where standard softmax classifiers degrade in skewed ways. Zhu et al. [3] and Lu and Steinerberger [4] provide theoretical analyses. ETF heads have since improved long-tailed recognition, incremental learning, and semi-supervised classification.

To our knowledge, no prior work extends ETF classifiers to explicit class hierarchies. Hierarchical losses based on label smoothing or tree-structured prototypes [5] encode hierarchy implicitly but do not exploit the geometric optimality of the ETF at each level.

3 Data

3.1 REMIA project corpus

Our study is based on the REMIA project dataset (Table 1), previously introduced in [8], which is composed of two complementary sources of imagery:

Real excavated sherds. These are physical sherds collected at Saran (Loiret, France) and digitized with an EinScan-SP 3D scanner. Wheel labels are provided by archaeologists when an authoritative association exists. Real sherds carry the natural variability described in Section 1: erosion, partial imprints, curvature, application defects.

Facsimiles. Following the experimental archaeology methodology of [8], new wooden wheels were carved by hand (mimicking the original tools) and rolled on flat strips of modeling paste; the strips were then digitized with the same EinScan-SP scanner. The resulting labels are secure by design (no ambiguity about which wheel produced the imprint). Facsimiles carry less natural variability, in particular no erosion and no curvature; nevertheless, they are produced under varying pressure levels to better approximate real sherds and partially reproduce surface degradation effects.

The two sources together yield 1,123 images grouped into 11 thumbwheels (sub-classes) and 4 typological super-classes:

- **Squares (type G)**: γ, G8 (179 images)
- **Squares (type H)**: η, H1, H12, H13 (322 images)
- **Chevrons (type L)**: λ, λ_1 (204 images)
- **Diamonds (type A)**: α, α_1, A2 (418 images)

Greek letters denote thumbwheels represented *only* by facsimiles (γ, η, λ, α); Latin letters denote thumbwheels containing real excavated sherds (G8, H1, H12, H13, A2).

One methodological choice distinguishes our corpus from [8]. To enlarge the dataset, Brahim et al. [8] split each scanned *facsimile* imprint into three subcrops treated as independent samples, reaching a total corpus of ∼2,500 images (real sherds and facsimiles combined). We instead keep each physical object as a single, indivisible sample and split train and test at the object level. This avoids

Table 1: REMIA project corpus: number of samples per thumbwheel (sub-class) and per source.

	Squares (G)		Squares (H)				Chevrons (L)		Diamonds (A)			Total
	γ	G8	η	H1	H12	H13	λ	λ_1	α	α_1	A2	
Real	0	30	0	104	20	48	0	0	0	0	203	405
Facsim.	149	0	150	0	0	0	50	154	165	50	0	718
Total	149	30	150	104	20	48	50	154	165	50	203	1123

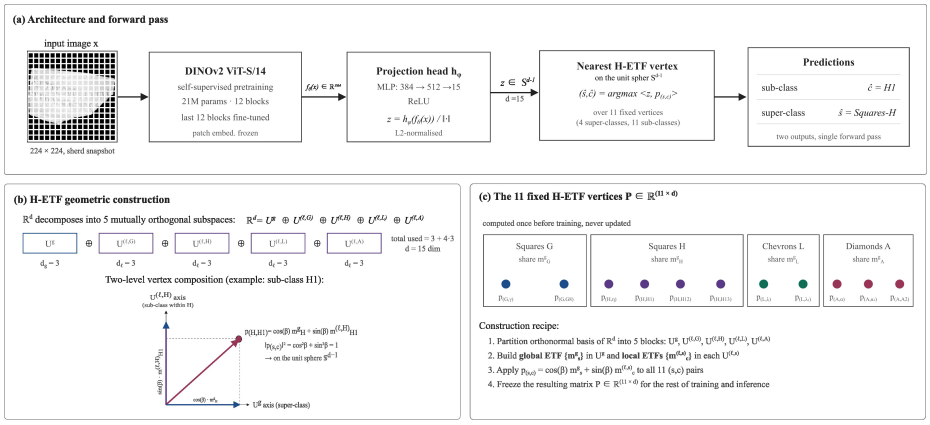


Fig. 2: Overview of H-ETF. **(a)** A DINOv2 ViT-S/14 backbone with partial fine-tuning produces a 384-dim feature, projected by a two-layer MLP to an L2-normalised embedding $z \in \mathbb{S}^{d-1}$ (d=15); classification reduces to nearest-vertex assignment among 11 fixed vertices, returning both sub- and super-class predictions in one forward pass. **(b)** Geometric construction: $\mathbb{R}^d$ decomposes into five mutually orthogonal subspaces, and each vertex is the orthogonal sum $p_{s,c} = \cos(\beta)\, m_s^g + \sin(\beta)\, m_c^{\ell,s}$, lying on the unit sphere. **(c)** The 11 vertices, grouped by super-class; the matrix $P \in \mathbb{R}^{11\times d}$ is computed once and frozen throughout training and inference.

within-object information leakage, since adjacent crops of the same imprint share illumination, scan artefacts, and partial pattern overlap.

3.2 Preprocessing

Each 3D scan is converted to a 2D shaded view that captures the ∼1–2 mm deep relief imprint and feeds the recognition pipeline. The preprocessing chain consists of two steps: (i) RANSAC fit of the dominant plane of the point cloud and rotation to align the decorated face with the image plane, and (ii) rendering of a 2D shaded snapshot. All snapshots are resized to 224×224 before being fed to the network.

4 Hierarchical ETF Recognition

The Hierarchical-ETF recognition pipeline (Fig. 2a) has three stages. A DINOv2 ViT-S/14 backbone [7] extracts a 384-dim feature $f_\theta(x)$; a two-layer MLP projects it to an L2-normalised embedding $z \in \mathbb{S}^{d-1}$ (d=15); classification is the nearest-vertex assignment among 11 fixed targets $\{p_{s,c}\}$ described below. The targets have no learnable parameters: they are computed once and frozen.

4.1 Background: Simplex ETF

A K-vertex simplex ETF in $\mathbb{R}^d$ ($d \geq K-1$) is a set of unit vectors $\{m_1, \ldots, m_K\}$ satisfying

$$\langle m_i, m_j \rangle \;=\; \begin{cases} 1 & i = j, \\ -\frac{1}{K-1} & i \neq j, \end{cases} \tag{1}$$

i.e. a regular simplex centred at the origin. This is the configuration that maximises the minimum pairwise angle and, as shown by [1, 2], the geometry to which classifier weights and class means converge under Neural Collapse. A flat 11-vertex ETF would treat every sub-class symmetrically and ignore the typological hierarchy. The next section shows how to preserve the ETF optimality at *both* hierarchy levels.

4.2 Hierarchical ETF Construction

Let $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_4\}$ be the four super-classes (G, H, L, A), with $\mathcal{S}_s$ containing n_s sub-classes (n_1=2, n_2=4, n_3=2, n_4=3; total 11). We decompose $\mathbb{R}^d$ into five mutually orthogonal subspaces (Fig. 2b):

- a *global* subspace U^g of dimension d_g=3, hosting a 4-vertex ETF $\{m_1^g, \ldots, m_4^g\}$ for the super-classes;
- four *local* subspaces $U^{\ell,1}, \ldots, U^{\ell,4}$, each of dimension d_ℓ=3, with $U^{\ell,s}$ hosting an n_s-vertex ETF $\{m_1^{\ell,s}, \ldots, m_{n_s}^{\ell,s}\}$ for the sub-classes of $\mathcal{S}_s$.

The five subspaces are obtained by partitioning a fixed orthonormal basis of $\mathbb{R}^d$ into consecutive blocks of sizes $d_g, d_\ell, d_\ell, d_\ell, d_\ell$, requiring $d \geq d_g + 4\,d_\ell = 15$.

The H-ETF vertex for sub-class c within super-class s is

$$p_{s,c} \;=\; \cos\beta\; m_s^g \;+\; \sin\beta\; m_c^{\ell,s}, \qquad \beta \in [0, \tfrac{\pi}{2}]. \tag{2}$$

Because the two components live in orthogonal subspaces, $\|p_{s,c}\|^2 = \cos^2\beta + \sin^2\beta = 1$: all eleven vertices lie on $\mathbb{S}^{d-1}$, matching the normalisation of z. The mixing angle β controls the hierarchy: $\beta \to 0$ merges sub-classes onto their super-class vertex; $\beta \to \pi/2$ recovers a flat ETF. We fix $\beta = \pi/4$ (equal weight to both levels). The full vertex matrix $P \in \mathbb{R}^{11 \times d}$ (Fig. 2c) is computed once before training.

Proposition 1 (Orthogonal decomposition). *Under* (2)*, the super-class and sub-class ETF structures are recovered exactly by orthogonal projection:* $P_{U^g}\, p_{s,c} = \cos\beta\; m_s^g$ *and* $P_{U^{\ell,s}}\, p_{s,c} = \sin\beta\; m_c^{\ell,s}$.

Proof. P_{U^g} fixes U^g and annihilates every $U^{\ell,s'}$; symmetrically, $P_{U^{\ell,s}}$ fixes $U^{\ell,s}$ and annihilates U^g and every $U^{\ell,s'}$ with $s' \neq s$.

Consequently, whenever the trained features align to the H-ETF targets, Neural Collapse [1, 2] is attained at both hierarchy levels simultaneously, with no trade-off between them.

4.3 Training

We fine-tune the last 12 transformer blocks and the projection head (the patch embedding and the vertex matrix P are frozen) with a hierarchical anchor loss:

$$\mathcal{L}(z) \;=\; \underbrace{\|z - p_{y_s,y_c}\|_2^2}_{\mathcal{L}_{\text{sub}}} + \lambda_s \underbrace{\|P_{U^g}\, z - \cos\beta\; m^g_{y_s}\|_2^2}_{\mathcal{L}_{\text{sup}}}, \qquad \lambda_s = 1. \tag{3}$$

$\mathcal{L}_{\text{sub}}$ pulls the embedding toward its target vertex. For unit-norm vectors, $\|z - p\|_2^2 = 2 - 2\langle z, p\rangle$, so $\mathcal{L}_{\text{sub}}$ is equivalent (up to a constant) to the dot-regression loss of [2]. $\mathcal{L}_{\text{sup}}$ anchors the super-class projection independently: because the super-class label is shared across many sub-classes and across both data sources (real and facsimile), it supplies a stronger gradient signal for sub-classes with few samples.

4.4 Inference

At test time, classification on the full hierarchy reduces to a single inner-product search:

$$(\hat{s}, \hat{c}) \;=\; \arg\max_{(s,c)} \langle z,\, p_{s,c}\rangle. \tag{4}$$

The same H-ETF supports two additional modes without any architectural change: (i) *super-class only*, $\hat{s} = \arg\max_s \langle P_{U^g}\, z,\, m^g_s\rangle$; (ii) *reduced-support*, where $p_{s,c}$ is replaced by the L2-normalised mean of z over a small support set and queries are classified by cosine similarity to the nearest prototype (see Section 5.4).

5 Experiments

5.1 Implementation details

Backbone: DINOv2 ViT-S/14 (21M parameters) loaded from the official Meta release. The last 12 transformer blocks are unfrozen for fine-tuning; the patch embedding is kept frozen. **Projection head**: 2-layer MLP $384 \to 512 \to 15$, with ReLU activation and L2-normalization at the output. **H-ETF geometry**: $d_g = 3$, $d_\ell = 3$ (since $\max_s(n_s - 1) = 3$ for Squares-H), $d = 15$, $\beta = \pi/4$. **Optimization**: AdamW with two parameter groups (head: lr $= 10^{-4}$, backbone: lr $= 10^{-5}$), weight decay 10^{-4}, cosine annealing over 100 epochs, batch size 32. Standard image augmentation (horizontal flip p=0.5, rotation $\pm 10°$, ImageNet normalization). **Reproducibility**: each experiment is run with three random seeds $\{4, 42, 4096\}$; we report mean $\pm$ standard deviation across seeds. **Hardware**: 1 NVIDIA RTX A4000 GPU with xFormers acceleration.

Table 2: Protocol P1 (combined source). Mean ± std over 3 seeds.

Method	Sub-Acc	Sup-Acc
EfficientNet-B5 + softmax	93.3 ± 0.6	98.9 ± 0.6
EfficientNet-B5 + H-ETF	95.4 ± 0.1	99.5 ± 0.3
DINOv2 + softmax	94.8 ± 0.2	98.9 ± 0.5
H-ETF (ours)	$\mathbf{96.9 \pm 0.4}$	$\mathbf{99.9 \pm 0.1}$

5.2 Baselines

We compare H-ETF against three baselines:

- **DINOv2 + softmax + CE (same backbone)**: same DINOv2 ViT-S/14 backbone with the same partial fine-tuning regime, but the H-ETF head is replaced by a linear classifier with cross-entropy loss. This isolates the contribution of the H-ETF geometry from that of the backbone and the fine-tuning recipe.
- **EfficientNet-B5 + softmax + CE (external backbone)**: an ImageNet-pretrained EfficientNet-B5 with a linear classifier and cross-entropy loss, fully fine-tuned. We use the B5 variant, which has more parameters (28M) than our DINOv2 ViT-S/14 backbone (21M), keeping the comparison conservative for H-ETF. This baseline tests whether a different backbone family can match H-ETF.
- **EfficientNet-B5 + H-ETF + $\mathcal{L}(z)$**: the same EfficientNet-B5 backbone fully fine-tuned, but with the H-ETF head and loss $\mathcal{L}(z)$ replacing the linear classifier with cross-entropy. This isolates the contribution of the H-ETF head from that of the backbone family.

5.3 Protocol P1 – Source isolation

For each source $\in \{\text{real}, \text{facsimile}, \text{combined}\}$, we train and test on the same source-restricted subset using a stratified 70/30 train/test split per sub-class. We report sub-class top-1 accuracy and super-class top-1 accuracy.

Table 2 reports the combined-corpus results, which is the headline configuration of the paper. Our H-ETF surpasses both same-backbone (DINOv2+softmax) and external (EfficientNet-B5+softmax) baselines on sub-class accuracy. The super-class accuracy is essentially saturated at 99.9%, confirming the hypothesis of Proposition 1: the orthogonal subspace decomposition imposes super-class separation as a structural property.

Table 3 reports the source-isolation breakdown. Two observations stand out. First, H-ETF on facsimiles alone reaches 99.4% accuracy, indicating that the controlled imagery is essentially saturated. Second, on real shards alone, the gap between H-ETF and the softmax baselines widens significantly: +3.1pp over the same-backbone baseline and +6.4pp over the external one. This is the regime where the geometric inductive bias of H-ETF compensates for data scarcity.

Table 3: Protocol P1 (source isolation). Sub-class accuracy, mean ± std over 3 seeds.

Method	Real	Facsimile
EfficientNet-B5 + softmax	84.0 ± 2.5	97.9 ± 0.6
EfficientNet-B5 + H-ETF	86.5 ± 1.4	98.5 ± 0.5
DINOv2 + softmax	87.3 ± 4.9	98.6 ± 0.9
H-ETF (ours)	$\mathbf{90.4 \pm 0.8}$	$\mathbf{99.4 \pm 0.7}$

Note on the Brahim et al. numbers. The reported numbers from [8] are not directly comparable to ours for two reasons. First, as detailed in Section 3, [8] split each facsimile imprint into three sub-crops treated as independent samples (∼2,500 total), whereas we keep each physical imprint as a single sample (1,123 total) and split at the object level to prevent within-object information leakage. Under our stricter protocol, an EfficientNet-B5 trained and evaluated under the same training and split protocol as H-ETF reaches 93.3% on the combined corpus and 84.0% on real shards. Second, [8] does not report multiple runs or seed-level variance; their tables are point estimates of unknown stability, whereas all our results are averaged over three seeds with explicit standard deviations. For reference, [8] reports 97.66% on the combined corpus and 81.97% on real shards alone with an EfficientNet-B5; our H-ETF reaches 90.4% on real shards (Table 3), an +8.4pp improvement over their published real-only figure, although the comparability caveats above apply. In all cases, the H-ETF results in this paper should be compared against our reproduction of EfficientNet-B5 (rows above), which controls for protocol differences.

5.4 Protocol P2 – Low-data evaluation on real shards

Motivation. Protocol P1 measures accuracy when the trained classifier head is used directly. It does not, however, probe the *geometry* of the learned embedding space itself: whether sub-classes form clusters compact and well-separated enough to be recovered from only a few exemplars. This is the property H-ETF is designed to induce, since aligning features to a fixed ETF drives Neural Collapse at both hierarchy levels. P2 tests it directly. We discard the trained classifier head and, for each $K \in \{1, 3, 5\}$, sample K shards per class as a *reduced support set*, compute class prototypes as the L2-normalized mean of their embeddings, and classify the remaining shards by cosine similarity to the nearest prototype. A method whose embedding space is genuinely well-structured should degrade gracefully as K shrinks; one that relies on a heavily parameterized decision boundary should not. P2 thus complements P1: P1 evaluates the classifier, P2 evaluates the representation. We run P2 on real shards, the regime where representation quality matters most.

Setup. For H-ETF, prototypes are computed in the H-ETF embedding space (i.e., the post-projection z) and re-normalized to lie on the unit sphere. For

Table 4: Protocol P2 (reduced-support evaluation on real shards, 4 sub-classes, H12 excluded). Mean ± std over 3 seeds with 50 draws per seed.

	Accuracy			Balanced accuracy		
Method	K=1	K=3	K=5	K=1	K=3	K=5
Eff.Net-B5 + softmax	82.2 ± 4.1	88.1 ± 2.8	89.5 ± 2.7	76.8 ± 4.2	83.6 ± 2.7	84.8 ± 3.0
DINOv2 + softmax	85.9 ± 0.4	91.0 ± 0.9	92.6 ± 0.5	79.1 ± 0.7	85.9 ± 1.0	87.9 ± 1.0
H-ETF (ours)	$\mathbf{92.5 \pm 1.3}$	$\mathbf{95.1 \pm 0.7}$	$\mathbf{95.7 \pm 0.9}$	$\mathbf{89.8 \pm 2.4}$	$\mathbf{92.8 \pm 1.3}$	$\mathbf{93.3 \pm 1.4}$

the two softmax baselines, prototypes are computed from the L2-normalized backbone features (pre-classifier) for direct comparability. We exclude H12 from this evaluation: with only ~ 6 test samples, the query set after sampling K=5 supports is too small for stable estimation. The remaining 4 sub-classes (G8, H1, H13, A2) are evaluated with $K \in \{1, 3, 5\}$ supports per class. For each (K, seed) pair we use 50 random support draws; we report the mean accuracy over draws, then aggregate mean ± std across the 3 seeds.

Note on terminology. P2 evaluates the discriminability of the embedding space over classes seen during training; it is not few-shot *generalization* to novel classes, which would require disjoint train/test classes and a different training protocol. We use "reduced-support" rather than "few-shot" throughout to mark this distinction.

Results. Table 4 reports both standard accuracy and balanced accuracy. H-ETF outperforms both baselines at every K and on both metrics. The largest gap appears at K=1 (+6.6pp over the better baseline), the regime where the pre-registered ETF geometry has the largest leverage: the prototype is computed from a single sample, and the inductive bias of the geometry compensates for the noise. As K grows, the gap narrows but H-ETF retains a clear advantage. The baselines' ranks are unstable across K: EfficientNet-B5+softmax beats DINOv2+softmax at K=1, but the order reverses at K=5.

5.5 Qualitative analyses

The row-normalized confusion matrices of Fig. 3 reveal where the three methods differ. All three reach 100% accuracy on the four facsimile-only sub-classes (γ, η, λ, α) and on the populous real sub-class A2. The Squares-H sub-family (H1, H12, H13) is where the methods diverge most. On H1, H-ETF reaches 0.84, ahead of EfficientNet-B5+softmax (0.81) and DINOv2+softmax (0.71); on H13, H-ETF reaches 0.93, against 0.79 for both baselines. H12, with only ~ 6 test samples, remains the residual bottleneck: H-ETF and DINOv2+softmax both resolve 50% of H12 samples correctly, whereas EfficientNet-B5+softmax collapses H12 entirely (0%, all samples absorbed by H1 and H13); the absolute number of samples is too low to draw a definitive conclusion. A further difference appears on

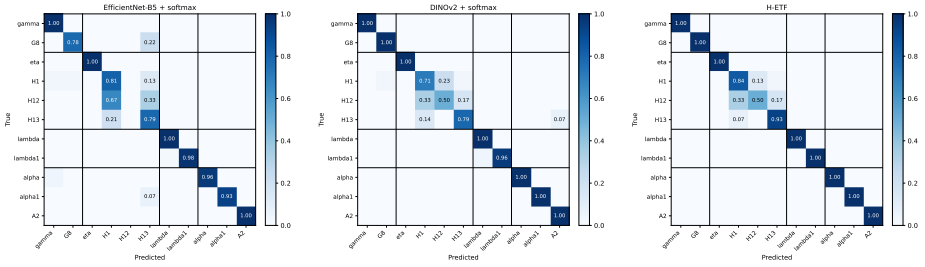
G8, which EfficientNet-B5+softmax classifies at only 0.78 (with 0.22 leaking to H13), while H-ETF and DINOv2+softmax both reach 1.00. Super-class confusion is essentially absent under H-ETF, in line with Proposition 1.

The PCA projections of Fig. 4 illustrate the structural property targeted by the H-ETF construction. Projecting the embedding z onto the super-class subspace U^g collapses each super-class to a tight cluster around the corresponding global ETF vertex (left panel). In the full embedding space, the super-class centroids remain spatially separated, while sub-classes within each super-class occupy mutually orthogonal directions around their parent centroid (right panel). This visual evidence aligns with the algebraic decomposition of Proposition 1: the global ETF geometry is recovered from the super-class projection, and subclass identity is encoded in the orthogonal complement.

A closer look confirms that, for almost every sub-class, the empirical centroid coincides with its theoretical H-ETF vertex—the visual signature of Neural Collapse, with the network placing features exactly where the fixed geometry prescribes. The only departure is the real Squares-H family. H1 and H13 do not collapse as tightly as the other sub-classes, but their centroids still sit relatively close to their theoretical vertices; H12, in contrast, drifts markedly away from its own vertex toward the H1 region. This is consistent with the confusion matrices of Fig. 3, where H12 is the dominant residual error, with a third of its samples assigned to H1. An imperfect collapse in embedding space and cross-confusion at the classifier level are two views of the same difficulty: separating wheels that differ only by a few incision lines and for which very few real samples are available.

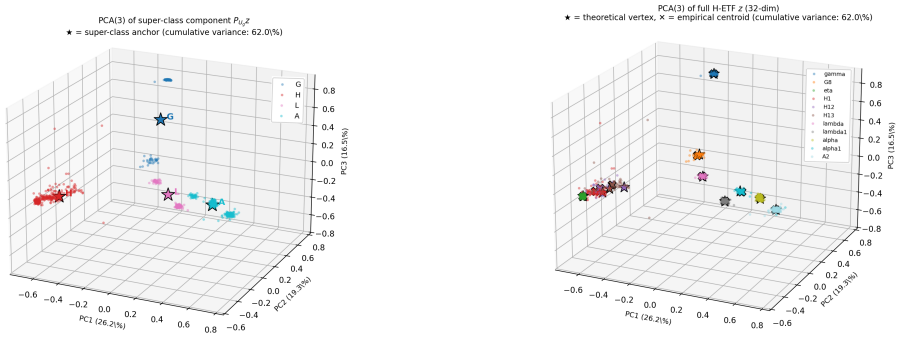
6 Discussion and Conclusion

We introduced the Hierarchical Equiangular Tight Frame (H-ETF), a fixed two-level simplex geometry that encodes typological super-classes and thumbwheel



(a) Eff.Net-B5 + softmax (b) DINOv2 + softmax (c) H-ETF (ours)

Fig. 3: Row-normalized confusion matrices on the P1/combined test set (seed 4096). Rows: true classes; columns: predictions; black lines delimit the four superclasses. All three methods recover super-classes nearly perfectly. Discriminative differences concentrate within the Squares-H sub-family (H1, H12, H13), where H-ETF reduces the cross-confusion seen in both softmax baselines.



(a) Super-class projection $P_{U^g}z$ (b) Full H-ETF embedding z

Fig. 4: 3D PCA projections of the H-ETF embedding on the combined test set. **(a)** Projecting z onto the super-class subspace U^g collapses samples to four well-separated clusters, one per super-class (G, H, L, A), each centered on its global ETF vertex. **(b)** In the full embedding space, sub-class clusters within each super-class are visible as orthogonal offsets from their parent centroid.

sub-classes in mutually orthogonal subspaces. Mounted on a DINOv2 ViT-S/14 backbone with partial fine-tuning, H-ETF reaches 96.9±0.4% sub-class accuracy and 99.9±0.1% super-class accuracy on the REMIA combined corpus, surpassing both same-backbone and external softmax baselines.

The advantage is twofold. First, the orthogonal decomposition imposes super-class separation as a structural property rather than a learned outcome: super-class accuracy stays at or above 99% across all configurations, so residual errors localize within typological families where they are easier to interpret and flag for archaeologist review. Second, the pre-registered geometry provides a strong inductive bias in low-data regimes (P1/real, P2/K=1), precisely where softmax classifiers are most exposed to overfitting on majority classes. The same-backbone ablation confirms that this gain stems from the head, not the backbone (+3.1 pp on real shards). We further note that the H-ETF head interacts more favorably with a self-supervised backbone (DINOv2) than with an ImageNet-supervised one (EfficientNet-B5), suggesting that the geometric constraint and self-supervised pre-training are complementary inductive biases.

Because the construction supports both direct classification and reduced-support nearest-prototype inference with no architectural change, it is directly applicable to field scenarios where only a handful of exemplars exist for a newly discovered wheel. We hope this contribution helps automate a long and tedious archaeological task, and that the H-ETF principle finds applications beyond ceramic pattern matching wherever a label hierarchy is available.

Acknowledgments The authors thank the Région Centre-Val de Loire (France) for funding the previous REMIA project, which produced the dataset used in this study.

Bibliography

[1] Papyan, Vardan, X. Y. Han, and David L. Donoho. "Prevalence of neural collapse during the terminal phase of deep learning training." Proceedings of the National Academy of Sciences 117.40 (2020): 24652-24663.

[2] Yang, Yibo, et al. "Inducing neural collapse in imbalanced learning: Do we really need a learnable classifier at the end of deep neural network?." Advances in Neural Information Processing Systems 35 (2022): 37991-38002.

[3] Zhu, Zhihui, et al. "A geometric analysis of neural collapse with unconstrained features." Advances in Neural Information Processing Systems 34 (2021): 29820-29834.

[4] Lu, Jianfeng, and Stefan Steinerberger. "Neural collapse under cross-entropy loss." Applied and Computational Harmonic Analysis 59 (2022): 224-241.

[5] Bertinetto, Luca, et al. "Making better mistakes: Leveraging class hierarchies with deep networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[6] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[7] Oquab, Maxime, et al. "Dinov2: Learning robust visual features without supervision." arXiv preprint arXiv:2304.07193 (2023).

[8] Brahim, Khawla, et al. "Facsimiles-based deep learning for matching relief-printed decorations on medieval ceramic sherds." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[9] Debroutelle, Teddy, et al. "Automatic classification of ceramic sherds with relief motifs." Journal of Electronic Imaging 26.2 (2017): 023010-023010.

[10] Chetouani, Aladine, et al. "Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling." Pattern Recognition Letters 131 (2020): 1-7.

[11] Chetouani, Aladine, et al. "Classification of ceramic shards based on convolutional neural network." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[12] Kampel, Martin, and Robert Sablatnig. "Color classification of archaeological fragments." Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. Vol. 4. IEEE, 2000.

[13] Smith, Patrick, et al. "Classification of archaeological ceramic fragments using texture and color descriptors." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, 2010.

[14] Makridis, Michael, and Petros Daras. "Automatic classification of archaeological pottery sherds." Journal on Computing and Cultural Heritage (JOCCH) 5.4 (2013): 1-21.

[15] Anichini, F., et al. & Zallocco, M.(2020). Developing the ArchAIDE application: a digital workflow for identifying, organising and sharing archaeo-

logical pottery using automated image recognition." Internet Archaeology 52 (2020).
[16] Anichini, Francesca, et al. "The automatic recognition of ceramics from only one photo: The ArchAIDE app." Journal of Archaeological Science: Reports 36 (2021): 102788.
[17] Tyukin, Ivan, et al. "Exploring automated pottery identification [Arch-I-Scan]." Internet Archaeology 50.10.11141 (2018).
[18] Parisotto, Simone, et al. "Unsupervised clustering of Roman potsherds via Variational Autoencoders." Journal of Archaeological Science 142 (2022): 105598.
[19] Pawlowicz, Leszek M., and Christian E. Downum. "Applications of deep learning to decorated ceramic typology and classification: A case study using Tusayan White Ware from Northeast Arizona." Journal of Archaeological Science 130 (2021): 105375.
[20] Benhabiles, Halim, and Hedi Tabia. "Convolutional neural network for pottery retrieval." Journal of Electronic Imaging 26.1 (2017): 011005-011005.