This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

A Hyperparameters and Augmentation

All models use a ConvNeXt-tiny backbone optimized with AdamW (head LR 10^{-3}, weight decay 10^{-4} on parameters of two or more dimensions only), a cosine schedule, batch size 32, and inverse-frequency cross-entropy weighting $w_c = (1/\text{freq}_c)^{0.8}$. The dense base model trains end-to-end; each dual-head model is warm-started from the matching-seed checkpoint, then freezes the backbone and Euclidean head (including BatchNorm) and trains only the hyperbolic trunk, with selection on the logit-averaged fused balanced accuracy. We use five seeds $\{42, \ldots, 46\}$; settings that differ between the two models are in Table 4.

The hyperbolic head embeds features in a 16-dimensional Lorentz space and classifies by distance to fixed prototypes, placed once per dataset by 2500 steps of the metric-guided optimization – see Section 3. As the frozen Euclidean head has no trainable upstream parameters, the objective reduces to the hyperbolic cross-entropy ($\alpha = 0$, $\beta = 1$) plus a radius penalty toward radius 3.0. Per dataset we select the geoopt curvature $k \in \{0.5, 1.0, 2.0\}$ (convention of Appendix B) and radius penalty $\in \{0.01, 0.05, 0.1\}$ on validation balanced accuracy; as this is essentially invariant across the grid, ties break on validation NLL.

Training images are augmented at 224×224. The dermoscopic datasets (ISIC 2019 and DERM12345) use random resized crops, full dihedral (D_4) flips and rotations, mild distortion, colour jitter, and coarse dropout; RetinaMNIST uses a conservative subset (mild crop, horizontal flip, small rotation, light brightness and contrast jitter) to preserve its small, colour-encoded lesions. All images use ImageNet normalization.

B Curvature, Decorrelation, and Fused NLL

We sweep the Lorentz curvature parameter k (geoopt convention; larger k is *flatter*, i.e. closer to Euclidean) over $\{0.2, 0.3, 0.5, 1.5, 2.0, 2.8, 3.5, 4.0\}$ for each of the 5 backbones on each dataset, holding the radius penalty at its validation-selected value, and record the fused NLL $\text{NLL}_{\text{fused}} = \text{NLL}\big(\text{softmax}\big(\frac{1}{2}(\mathbf{z}^{\text{euc}} + \mathbf{z}^{\text{hyp}})\big), y\big)$, the Euclidean head's stand-alone NLL, NLL_{euc}, and the inter-head correlation.

Within each backbone, inter-head correlation rises with k (the hyperbolic head flattens toward the Euclidean one): the per-backbone correlation between

Table 4. Settings that differ between the dense Euclidean base model and the dual-head model; shared settings are given in the text.

	Dense Euclidean	Dual-head (Euclidean/hyperbolic)
Backbone LR	10^{-5}	frozen
Max epochs	100	50
Early-stop patience	20	10
Selection metric	val. bal. acc.	val. bal. acc. (fused)
Embedding dim	n/a	16
Target radius	n/a	3.0

k and inter-head correlation is positive on 13/15 backbones (ISIC 5/5, DERM 5/5, Retina 3/5). Since geoopt k runs inverse to the curvature axis of Figure 3 (higher means less Euclidean), this matches the negative curvature relationship in the main text: the heads decorrelate as the hyperbolic head becomes less Euclidean. With k mean-centred ($\tilde{k}$), a linear mixed model of $\mathrm{NLL}_{\mathrm{fused}}$ on $\tilde{k}$ (dataset fixed effects, a curvature-by-dataset interaction, and per-backbone random slopes) gives a positive pooled slope, $\hat{\beta}_1 = 0.016$ (95% CI $[0.004, 0.028]$, $p = 0.02$), but a significant curvature-by-dataset interaction ($p = 1.2 \times 10^{-5}$): the effect is carried by ISIC and absent on DERM. Notably, DERM is also the dataset whose Euclidean head is strongest ($\mathrm{NLL}_{\mathrm{euc}} \approx 0.5$ to 0.7, vs. 1.1 to 1.9 elsewhere), suggesting curvature helps only where the Euclidean head leaves headroom, consistent with ensemble theory in which member decorrelation *bounds* rather than determines performance. We report this cautiously as a between-dataset association, since $\mathrm{NLL}_{\mathrm{euc}}$ is collinear with dataset and is itself a component of $\mathrm{NLL}_{\mathrm{fused}}$.

Table 5. Fusion performance as maximum validation balanced accuracy on ISIC 2019

	Mixed	Dual Euclidean
Logit Fusion	0.897 ± 0.038	0.891 ± 0.041
Prob Fusion	0.886 ± 0.037	0.887 ± 0.039