Fibers of Asymmetric Similarity: A Framework for Clinical and Imaging Data

Johanna P. Müller[1], Matthew Baugh[2], Robert Wright[5], Thomas G. Day[3,4,5], Reza Razavi[5], and Bernhard Kainz[1,2,3]

[1] Dept. AIBE, Friedrich-Alexander University Erlangen–Nürnberg, DE
johanna.paula.mueller@fau.de
[2] Dept. Computing, Imperial College London, UK
[3] School of Biomedical Engineering & Imaging Sciences, King's College London, UK
[4] Evelina London Children's Healthcare, GSST NHS Foundation Trust, UK
[5] Fraiya Ltd., London, UK

Abstract. Prenatal screening for congenital heart disease (CHD) is challenged by rare pathology, heterogeneous data, uneven modality contributions, and frequent missing measurements. We propose a unified asymmetric-similarity methodology: a single Tversky prototype-comparison principle applied independently to structured clinical data and to medical imaging, rather than a fused multimodal model. For clinical data, a composite loss combining Tversky-prototype attraction, Tversky-contrastive separation, and entropy-regularised soft assignment learns embeddings capturing graded diagnostic similarity under severe class imbalance. For imaging, similarity fields classify by comparing frozen encoder features to learned prototypes via asymmetric Tversky measures, defining class membership through fiber inclusion in the similarity geometry. On the CARDIUM dataset (7.19 % CHD prevalence), our schema-aware Tversky loss raises F1 from 0.294 to 0.562 with perfect recall using only clinical data, while similarity fields improve F1 by 2-4× over identical frozen backbones. On thyroid nodule classification (Stanford AIMI TUC), hypergraph Tversky embeddings achieve AUC 0.81 from relational TI-RADS features alone. By replacing symmetric distances with asymmetric prototype comparisons, the framework offers a principled solution for learning under class imbalance, and missing data. Code is available at https://github.com/ividja/tversky-fibers.

Keywords: Ultrasound screening · Class imbalance · Similarity fields.

1 Introduction

Congenital heart disease (CHD) is the most common birth defect, affecting approximately 1 % of live births [6], yet prenatal detection rates remain highly variable, ranging from 30 % in routine obstetric screening to around 90 % with specialised fetal echocardiography [3,1]. This gap reflects challenges of ultrasound screening: image quality varies with operator, fetal position, and maternal habitus; the pathology is rare even in referred populations (∼7 %) [3]; and

decisions must integrate imaging with often-incomplete maternal risk factors. Such settings expose a limitation of standard objectives: binary cross-entropy and symmetric distances treat all errors equally, yet a missed CHD case is far costlier than a false alarm, while excessive false positives still burden limited resources through unnecessary referrals. Moreover, comparing a rare pathology to a common normal is not the same as the reverse: similarity is inherently *asymmetric* [20].

Tversky's contrast model formalises this: similarity weighs shared against distinctive features, so $S(A, B) \neq S(B, A)$ in general [20,21], giving explicit control over missed versus spurious features and thus over the precision-recall trade-off [15,14,2]. Whereas Tversky losses have served image-space segmentation [15,23], we apply asymmetric similarity in latent space to shape embeddings and prototype comparisons. Prototype-based classification [16] is a natural vehicle: the model learns *what a class looks like* and classifies by similarity rather than a discriminative boundary. Unlike prototypical networks using symmetric Euclidean or cosine distances, we use asymmetric Tversky similarity for prototypes in both structured clinical data and imaging. Our contributions are:

1. **Tversky similarity for clinical data embeddings.** A composite loss combining bidirectional Tversky-prototype attraction, Tversky-contrastive separation, and entropy-regularised soft prototype assignment that learns embeddings robust to class imbalance and missing data. We introduce schema-aware and hypergraph encoders tailored to independent and relational clinical variables, respectively.
2. **Similarity fields for image classification.** We formalise image-level and patient-level classification through similarity fields [11]: frozen encoder features are compared to learned prototypes via an asymmetric Tversky function, with class membership defined as inclusion in the resulting similarity fiber. This requires *no fine-tuning* of the image encoder, enabling plug-and-play use of any pretrained backbone.

Related Work. Prototype-based learning from few-shot metric learning [16] is widely used in medical segmentation, with prototypes refined via contrastive or clustering strategies [17,19,7,8]. The Tversky index generalises Dice and Jaccard by asymmetrically weighting false positives and negatives [15], and differentiable layers enable end-to-end similarity learning [4]. Unlike prior Tversky use on segmentation masks or image-feature contrastive objectives [15,23], we operate in latent space with learned prototypes, bidirectional asymmetry, and explicit margins, for both clinical data and frozen image features. Foundation models (DINOv2 [12], BiomedCLIP [25], CLIP [13]) supply rich features without fine-tuning, which our similarity fields exploit directly for rare-disease settings.

2 Method

2.1 Tversky Similarity for Clinical Data

Encoders. Clinical data here are structured tabular features (numeric, ordinal, categorical); patient features $x \in \mathbb{R}^d$ follow a schema encoding types, valid

ranges, ordinal orderings, and logical dependencies. We propose two encoder architectures matched to data structure. (1) A *schema-aware encoder* for independent clinical parameters (*e.g.*, maternal age, blood pressure, BMI) normalises numeric features as $\tilde{x}_i = (x_i - \mu_i)/\sigma_i$ and penalises out-of-range deviations via $P = \frac{1}{d}\sum_{i=1}^{d} w_i[\max(0, x_i - x_i^{\max}) + \max(0, x_i^{\min} - x_i)]$. Temporal or repeated measurements (*e.g.*, gestational age) $x \in \mathbb{R}^{T\times d}$ are embedded and pooled as $h = \sum_{t=1}^{T} f_{\text{clin}}(\tilde{x}_t)$, where $f_{\text{clin}} : \mathbb{R}^d \to \mathbb{R}^D$ maps heterogeneous features into a D-dimensional embedding space. (2) A *hypergraph encoder* for relational or interdependent features (*e.g.*, TI-RADS composition, echogenicity, and margins) embeds each variable via a feed-forward network with type embedding, prepends a patient token, and processes the sequence with a transformer to capture higher-order interactions, with a feed-forward fallback encoder for patients with few valid features.

Composite Tversky loss. Let $\{p_k\}_{k=1}^{K} \subset \mathbb{R}^D$ be learnable class prototypes. The asymmetric Tversky similarity $S_\alpha(a, b)$ is the Eq. 3 kernel on sigmoid-mapped embeddings $h, p \in (0, 1)^D$; its weights α, β are fixed (with $\beta > \alpha$ to prioritise recall), and only the projector and prototypes $\{p_k\}$ are learned. We define a bidirectional operator controlled by $\eta \in [0, 1]$: $\mathcal{S}_\alpha^\eta(h, p) = \eta\, S_\alpha(h, p) + (1-\eta)\, S_\alpha(p, h)$.

Tversky-Prototype loss $\mathcal{L}_{\text{Proto}}$. Soft assignment weights $w_{ik} = \exp(S_\alpha(h_i, p_k)/\tau)/\sum_j \exp(S_\alpha(h_i, p_j)/\tau)$ distribute each positive embedding h_i across prototypes with temperature τ. The loss attracts positive embeddings to their assigned prototypes, repels negatives beyond margin m, and regularises assignment entropy:

$$\mathcal{L}_{\text{Proto}} = \underbrace{\frac{1}{N_+}\sum_{i\in\mathcal{P}}\Big(1 - \textstyle\sum_k w_{ik}\,\mathcal{S}_\alpha^\eta(h_i, p_k)\Big)}_{\text{positive attraction}} + \underbrace{\frac{1}{N_-}\sum_{i\in\mathcal{N}}\max_k\big(S_\alpha(h_i, p_k) - m\big)_+}_{\text{negative repulsion}} - \underbrace{\frac{\lambda_{\text{ent}}}{N_+}\sum_{i\in\mathcal{P}}\textstyle\sum_k w_{ik}\log w_{ik}}_{\text{entropy reg.}}, \tag{1}$$

where $\mathcal{P}, \mathcal{N}$ denote positive and negative sets of size N_+, N_-.

Tversky-Contrastive loss $\mathcal{L}_{\text{Contr}}$ enforces intra-class cohesion and inter-class separation directly between embeddings:

$$\mathcal{L}_{\text{Contr}} = \frac{1}{\binom{N_+}{2}}\sum_{\substack{i,j\in\mathcal{P}\\ i<j}}\Big(1 - \mathcal{S}_\alpha^\eta(h_i, h_j)\Big) + \frac{2}{N_+N_-}\sum_{\substack{i\in\mathcal{P}\\ j\in\mathcal{N}}}\big(S_\alpha(h_i, h_j) - m\big)_+. \tag{2}$$

Cosine contrastive loss $\mathcal{L}_{\text{Cos}}$. A standard cosine-similarity contrastive term provides symmetric regularisation and can be scaled by class prevalence for imbalanced datasets. The full training objective (Fig. 1) combines three terms as $\mathcal{L} = \gamma_1\mathcal{L}_{\text{Proto}} + \gamma_2\mathcal{L}_{\text{Contr}} + \gamma_3\mathcal{L}_{\text{Cos}}$.

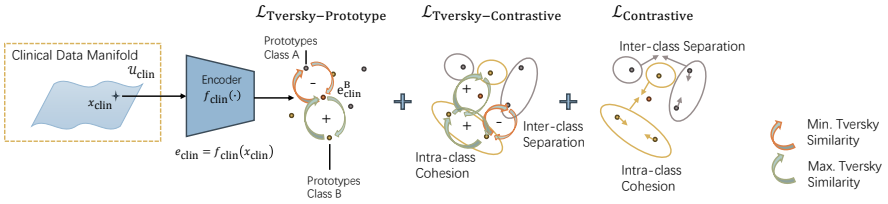

Fig. 1: **Tversky Similarity for Clinical Data.** Composite loss: $\mathcal{L}_{\text{Proto}}$ aligns embeddings to prototypes (bidirectional Tversky, soft assignment); $\mathcal{L}_{\text{Contr}}$ enforces intra-class cohesion and margin separation; $\mathcal{L}_{\text{Cos}}$ adds cosine regularisation. Temperature, entropy regularisation, and optional prevalence weighting improve robustness to imbalance.

2.2 Similarity Fields for Image Classification

We recast image classification with frozen features as *similarity-fiber* membership, eliminating backbone fine-tuning.

Similarity fields and fibers. Following Similarity Field Theory [11], let $\mathcal{U}$ be a universe with similarity field $S : \mathcal{U} \times \mathcal{U} \to [0, 1]$ satisfying reflexivity ($S(x, x) = 1$); neither symmetry nor transitivity is required, permitting the asymmetric, non-metric structure suited to imaging. A *concept* is an entity $c \in \mathcal{U}$, with *fiber* $F_\tau(c) = \{\, x \in \mathcal{U} \mid S(x, c) \geq \tau \,\}$ at threshold τ. Membership is graded by how far similarity to c exceeds τ (Fig. 2).

Embedding and prototype-parameterised concepts. A projector f_{img} maps a frozen encoder's output x_{img} (from input x'_{img}) into the embedding space, $e = f_{\text{img}}(x_{\text{img}}) \in [0, 1]^D$. For each class $k \in \{1, \ldots, C\}$, a set of M learnable prototype entities $\{P_k^{(j)}\}_{j=1}^{M} \subset [0, 1]^D$ parameterises the class concept. The similarity field is instantiated via a Tversky-type kernel:

$$S(e,\, p) \;=\; \frac{I(e,p)}{\max\!\big(\, I(e,p) + \alpha\, A(e,p) + \beta\, A(p,e),\; \varepsilon\big)}, \tag{3}$$

where $I(e,p) = \sum_i \min(e_i,\, p_i)$, and $A(e,p) = \sum_i (e_i - p_i)_+$, $\alpha, \beta \geq 0$ control the asymmetric penalties on features present in the embedding but absent in the prototype (α) and vice versa (β), and $\varepsilon > 0$ guards only against a zero denominator. When $e = p \neq \mathbf{0}$, the max is inactive and $S(e, e) = 1$ exactly; $S(\mathbf{0}, \mathbf{0}) := 1$ by convention, satisfying reflexivity for all $e \in [0, 1]^D$ (Def. 4, [11]). Class-wise similarity aggregates over prototypes as $s_k(e) = \max_j\, S(e,\, P_k^{(j)})$.

Multiple Instance Learning. As each patient has a variable number of images, we adopt multiple-instance learning (MIL): a patient is a *bag* of N images with embeddings $\{e_n\}_{n=1}^{N}$. A gated attention pooler produces a patient-level embedding

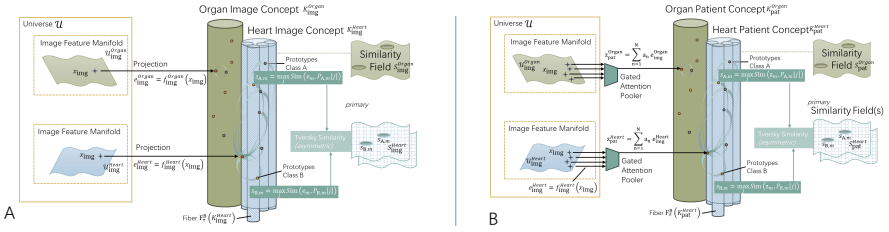

Fig. 2: **Similarity Fields for Image Classification.** A frozen encoder maps images into universe $\mathcal{U}$; prototypes $P_k^{(j)}$ parameterise each class concept K_k. The Tversky field (Eq. 3) gives each image (A) or patient (B) embedding graded membership in each class fiber. Classification follows from Eq. 5 from the fused decision field, with training, optimising the projection and prototype positions.

$z_{\text{pat}} = \sum_{n=1}^{N} a_n e_n$, with

$$a_n = \frac{\exp\big(\mathbf{w}^\top(\tanh(\mathbf{V}e_n) \odot \sigma(\mathbf{U}e_n))\big)}{\sum_{m=1}^{N} \exp\big(\mathbf{w}^\top(\tanh(\mathbf{V}e_m) \odot \sigma(\mathbf{U}e_m))\big)}, \tag{4}$$

where $\mathbf{V}, \mathbf{U} \in \mathbb{R}^{L \times D}$ and $\mathbf{w} \in \mathbb{R}^L$ are learnable attention parameters, and L is the attention hidden dimension. The pooled embedding z_{pat} replaces e in the similarity field, ensuring patient-level supervision.

Training objective. Classification is formulated as membership in a class-specific similarity fiber. Per-class scores are fused via convex combination, $s_C^{\text{fused}} = \sum_{k=1}^{C} w_k\, s_k$ with $w_k \geq 0$ and $\sum_k w_k = 1$, which preserves the $[0, 1]$ range and the similarity-field structure by Prop. 5 of SFT [11]. Inter-class separation is enforced with a multi-class hinge loss,

$$\ell(x) \;=\; \max\Big(0,\; \zeta - s_y(e) + \max_{k \neq y} s_k(e)\Big), \tag{5}$$

where $\zeta > 0$ is a margin. Only the prototypes (and the per-class convex weights w_k) are optimised via a single linear layer; the encoder and the asymmetry weights α, β remain frozen. The framework is lightweight and backbone-agnostic, directly applicable to encoders such as DINOv2 [12], BiomedCLIP [25], and CLIP [13].

3 Experiments

Datasets. We use two public ultrasound datasets. **CARDIUM** [22] targets prenatal CHD detection from 1,103 Colombian pregnancies (2013-2024; prevalence 7.19 %): 6,558 images across four cardiac views (GE Voluson E6/E8/E10) and 26 maternal variables. **Stanford AIMI Thyroid Ultrasound Cine-clip**

(TUC) [18] has 17,412 frames from 167 patients (192 biopsy-confirmed nodules) with TI-RADS descriptors, lesion metadata, and histopathology.

Implementation. Clinical embeddings use D=256, K=9 prototypes; image fields use M=5 prototypes per frozen backbone (DINOv2 [12], BiomedCLIP [25], CLIP [13], MedViT [9]. We use *patient-level* cross-validation (CARDIUM 3-fold by patient ID; TUC 2-fold stratified on the biopsy-confirmed annotations) at threshold 0.5, so no patient spans train and test. Tversky weights are fixed with $\beta{>}\alpha$ for recall: α=0.3, β=0.7 (clinical) and α=0.2, β=0.8 (image), with m=0.3, τ=0.1; the clinical loss reuses the Eq. 3 kernel on sigmoid-mapped embeddings $h, p \in (0,1)^D$, so $\beta{>}\alpha$ penalises missing target-class features more heavily than extra embedding features (recall-prioritised rule-out), learning only the projector, prototypes, and convex weights. MIL draws N=4 images per patient per epoch; this variance is captured over n=5 seeds ($\pm$0.073, MedViT HF), smaller than the gap to the fine-tuned baseline (0.519 vs. 0.345). Prototypes are Xavier-initialised with a 5/3-epoch LR warm-up (MIL/single-instance); remaining grid-searched hyperparameters are in the released code, and the Tversky kernel adds one element-wise op over cosine ($\approx$ 5% wall-clock on A100).

Clinical Data Embeddings. Tab. 1 reports CARDIUM results. The schema-aware encoder with the full Tversky loss performs best and exhibits the lowest variance, indicating stable optimisation. On TUC, the hypergraph encoder with $\mathcal{L}_{\text{Contr}}$ reaches AUC $0.81_{\pm.11}$ and accuracy $0.84_{\pm.09}$ from structured features alone, matching CNN-based SOTA [24]. Encoder choice should match data structure: schema-aware for independent variables (CARDIUM), hypergraph for relational features (TI-RADS; Fig. 3).

Image Classification via Similarity Fields On CARDIUM (Fig. 4), a parameter-matched baseline (frozen linear probe, same learnable-parameter budget as our prototype head) reaches high AUC (up to 0.830) but very low CHD-class F1 (0.106-0.238), reflecting majority-class bias under severe imbalance. Replacing *only* this head with learned prototypes and our Tversky similarity field substantially improves F1 across all backbones, e.g., BiomedCLIP 0.118 $\rightarrow$ **0.440**, DINOv2 0.106 $\rightarrow$ 0.378, and MedViT HF 0.126 $\rightarrow$ 0.347, without any encoder fine-tuning. In the multi-instance setting, MedViT HF achieves the highest F1 $\mathbf{0.519}_{\pm.073}$ and BiomedCLIP the highest precision $\mathbf{0.587}_{\pm.121}$, optimising only $C \times M \times D$ prototype parameters and one projection layer. Our approach outperforms the reproduced fine-tuned MedViT baseline on the CHD class (0.519 vs. 0.345) and achieves AUC 0.786-0.830 from 6,558 images at 7.19 % prevalence, comparable to prior balanced studies with over 50,000 images [5].

Ablation Study We ablate on CARDIUM clinical data with the best configuration (schema-aware, D=256, K=9, n=5). Removing any term degrades F1 and AUC (Tab. 2, top), with $\mathcal{L}_{\text{Proto}}$ the largest contributor and $\mathcal{L}_{\text{Contr}}$ chiefly aiding precision. Reducing the bidirectional operator to a symmetric average (η=0.5) lowers F1 from 0.562 to 0.422 and inflates its standard deviation from $\pm$0.008 to $\pm$0.198 ($\sim$25$\times$): asymmetry both improves accuracy and stabilises optimisation under imbalance, accounting for the low variance of the full model. Unlike sample/output-space remedies (focal loss, SMOTE, cost-sensitive reweighting),

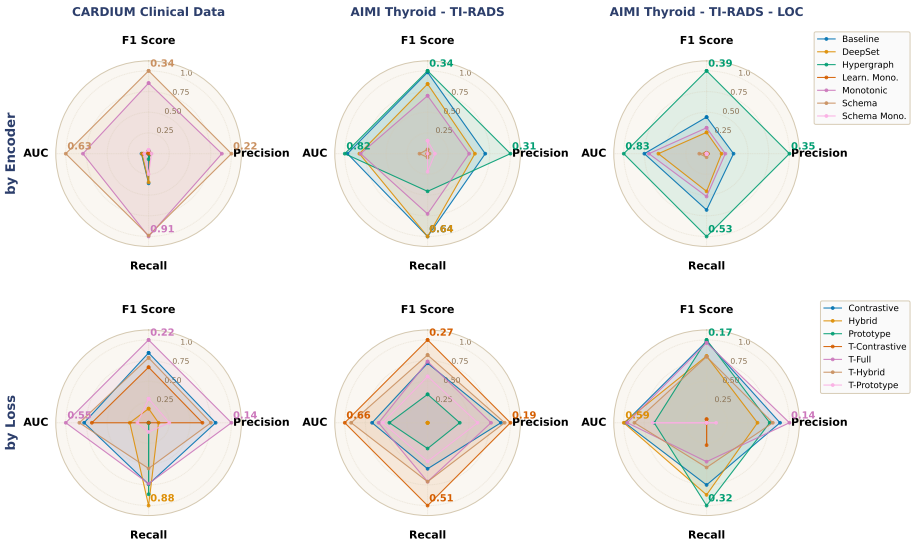

Fig. 3: **Encoder-loss interaction across datasets.** CARDIUM (left): independent variables, schema-aware encoders best. TUC TI-RADS (middle): relational, hypergraph encoders excel, especially with nodule location and size (right). Values averaged over prototype counts $[3, 6, 9]$ and dimensions $[256, 512]$; axes normalised.

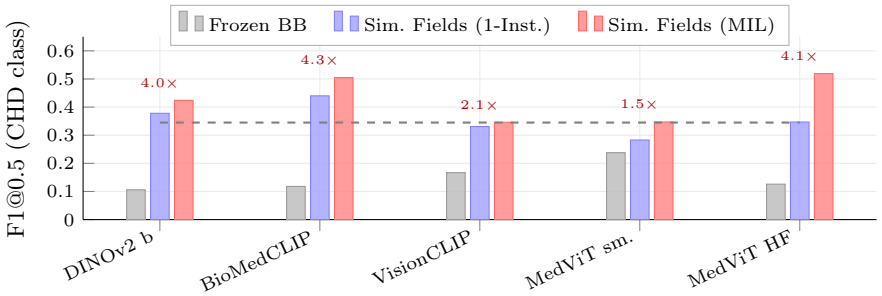

Fig. 4: **Similarity fields vs. frozen backbones on CARDIUM** (CHD-class F1 at threshold of 0.5). Grey: frozen features, parameter-matched linear head. Blue/red: similarity fields with learned prototypes (no fine-tuning), MIL gains (×) annotated. Dashed: fine-tuned MedViT baseline (0.345) [22]. Best (MedViT HF, MIL) improves +50 %, training only projection and prototypes.

Tversky asymmetry acts in embedding-prototype geometry; the symmetric row and prevalence-weighted cosine variants are its in-family analogues and are dominated at $p < 0.05$ (Fig. 5).

Discussion. The schema-aware Tversky encoder ranks first by critical-difference analysis (F1 0.562, ±0.008). Similarity fields improve minority-class F1 by 2-4×

Table 1: **Clinical data embeddings on CARDIUM** for CHD detection. Results averaged over n=5 runs (D=256, K=9 prototypes). **Best** overall; second-best per encoder group.

	Method	F1	Precision	Recall	AUC
	Vega et al. [22]	$0.294_{\pm.019}$	$0.192_{\pm.019}$	$0.634_{\pm.049}$	$0.794_{\pm.028}$
Monot.	Contrastive	$0.440_{\pm.175}$	$0.294_{\pm.130}$	$\mathbf{1.000}_{\pm.000}$	$0.733_{\pm.350}$
Monot.	Tversky-hybrid	$0.376_{\pm.189}$	$0.245_{\pm.145}$	$\mathbf{1.000}_{\pm.000}$	$0.738_{\pm.270}$
Monot.	Tversky-full	$0.335_{\pm.209}$	$0.221_{\pm.155}$	$0.920_{\pm.179}$	$0.554_{\pm.411}$
Schema	Tversky-full	$\mathbf{0.562}_{\pm.008}$	$\mathbf{0.391}_{\pm.007}$	$\mathbf{1.000}_{\pm.000}$	$\mathbf{0.900}_{\pm.003}$
Schema	Contrastive	$0.422_{\pm.198}$	$0.284_{\pm.150}$	$0.960_{\pm.089}$	$0.765_{\pm.238}$
Schema	Tversky-prototype	$0.391_{\pm.109}$	$0.278_{\pm.066}$	$0.680_{\pm.273}$	$0.809_{\pm.085}$

Table 2: **Ablation study on CARDIUM clinical data** (schema-aware encoder, D=256, K=9, n=5). Top: loss component ablation. Bottom: symmetric vs. asymmetric.

Configuration	F1	Precision	Recall	AUC
Full ($\gamma_1, \gamma_2, \gamma_3 > 0$)	$\mathbf{0.562}_{\pm.008}$	$\mathbf{0.391}_{\pm.007}$	$\mathbf{1.000}_{\pm.000}$	$\mathbf{0.900}_{\pm.003}$
w/o $\mathcal{L}_{\text{Proto}}$ (γ_1=0)	$0.264_{\pm.219}$	$0.171_{\pm.156}$	$0.760_{\pm.434}$	$0.591_{\pm.275}$
w/o $\mathcal{L}_{\text{Contr}}$ (γ_2=0)	$0.303_{\pm.211}$	$0.196_{\pm.153}$	$0.840_{\pm.261}$	$0.634_{\pm.263}$
w/o $\mathcal{L}_{\text{Cos}}$ (γ_3=0)	$0.264_{\pm.219}$	$0.171_{\pm.156}$	$0.760_{\pm.434}$	$0.591_{\pm.275}$
Symmetric (η=0.5)	$0.422_{\pm.198}$	$0.284_{\pm.150}$	$0.960_{\pm.089}$	$0.765_{\pm.238}$
Asymmetric (η=η^*)	$\mathbf{0.562}_{\pm.008}$	$\mathbf{0.391}_{\pm.007}$	$\mathbf{1.000}_{\pm.000}$	$\mathbf{0.900}_{\pm.003}$

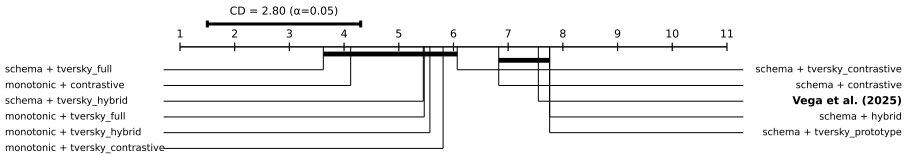

Fig. 5: **Critical difference diagram** (Nemenyi, α=0.05) ranks encoder-loss combinations by F1 (lower rank is better); connected methods are not significantly different. Schema+Tversky-full attains the best mean rank (≈3.6) and is significantly superior to methods ranked beyond ≈6.2.

over standard heads and beat fine-tuning while learning only prototype and projector parameters, the MIL variant reaching AUC 0.786-0.830 from 6,558 images. At 7.19 % prevalence with recall 1.0, precision 0.391 keeps referral burden tractable while far exceeding routine screening sensitivity ($\sim$ 30 % [3,1]); high recall is the screening-appropriate operating point. *Limitations:* prospective validation, scanner/domain shift, calibration, and referral burden at maximal recall remain open.

4 Conclusion

We presented asymmetric Tversky similarity learning for both structured clinical data and frozen-encoder imaging in rare-disease screening. On prenatal CHD (CARDIUM, 7.19 % prevalence) it improves clinical-data F1 by 91 % with perfect recall and lifts minority-class image F1 by 50 % over a fine-tuned baseline, with thyroid (TUC) results confirming cross-domain generalisability. Replacing symmetric distances with asymmetric prototype comparisons yields a principled, lightweight approach to the class imbalance and data heterogeneity of clinical screening.

Acknowledgments. We acknowledge the HPC resources provided by NHR@FAU under the NHR projects b143dc and b180dc. NHR funding is provided by federal

and Bavarian state authorities. NHR@FAU hardware is partially funded by the DFG - 440719683. Additional support was received by the ERC - project MIA-NORMAL 101083647, CHARMS 101246053, and ORACLE 101326871, DFG 513220538, 512819079, and by the state of Bavaria (HTA and the Bavarian Foundation Model imitative). We also acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation, and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023] [10]. We used coding agents and LLMs from Anthropic, OpenAI, Google, and Mistral AI, as writing aid, coding, experiment orchestration, and cluster monitoring.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bottelli, L., Franzè, V., Tuo, G., et al.: Prenatal detection of congenital heart disease at 12–13 gestational weeks: detailed analysis of false-negative cases. Ultrasound in Obstetrics & Gynecology **61**(5), 577–586 (2023). https://doi.org/10.1002/uog.26094
2. Day, T.G., Matthew, J., Budd, S.F., et al.: Ai to assist in the fetal anomaly ultrasound scan: A randomized controlled trial. NEJM AI **2**(4), AIoa2400747 (2025). https://doi.org/10.1056/AIoa2400747
3. Donofrio, M.T., Moon-Grady, A.J., Hornberger, L.K., et al.: Diagnosis and treatment of fetal cardiac disease: a scientific statement from the American Heart Association. Circulation **129**(21), 2183–2242 (2014). https://doi.org/10.1161/01.cir.0000437597.44550.5d
4. Doumbouya, M.K.B., Jurafsky, D., Manning, C.D.: Tversky neural networks: Psychologically plausible deep learning with differentiable tversky similarity. arXiv:2506.11035 (2025)
5. Gong, Y., Zhang, Y., Zhu, H., et al.: Fetal congenital heart disease echocardiogram screening based on dgacnn: Adversarial one-class classification combined with video transfer learning. IEEE Transactions on Medical Imaging **39**(4), 1206–1222 (2020). https://doi.org/10.1109/TMI.2019.2946059
6. Hoffman, J.I.E., Kaplan, S.: The incidence of congenital heart disease. Journal of the American College of Cardiology **39**(12), 1890–1900 (2002). https://doi.org/10.1016/S0735-1097(02)01886-7
7. Huang, W., Hu, J., Xiao, J., et al.: Prototype-guided graph reasoning network for few-shot medical image segmentation. IEEE Transactions on Medical Imaging (2024). https://doi.org/10.1109/TMI.2024.3459943
8. Li, Y., Daho, M.E.H., Conze, P.H., et al.: A review of deep learning-based information fusion techniques for multimodal medical image classification. Computers in Biology and Medicine **177**, 108635 (2024). https://doi.org/10.1016/j.compbiomed.2024.108635

9. Manzari, O.N., Ahmadabadi, H., et al.: Medvit: a robust vision transformer for generalized medical image classification. Computers in Biology and Medicine **157**, 106791 (2023). https://doi.org/10.1016/j.compbiomed.2023.106791
10. McIntosh-Smith, S., Alam, S.R., Woods, C.: Isambard-AI: a leadership class supercomputer optimised specifically for artificial intelligence (2024)
11. Ng, K.S.: Similarity field theory: A mathematical framework for intelligence. arXiv preprint arXiv:2509.18218 (2025)
12. Oquab, M., Darcet, T., Moutakanni, T., et al.: DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research (2024), https://openreview.net/forum?id=a68SUt6zFt
13. Radford, A., Kim, J.W., Hallacy, C., et al.: Learning transferable visual models from natural language supervision. In: Meila, M., Zhang, T. (eds.) ICML. PMLR, vol. 139, pp. 8748–8763. PMLR (18–24 Jul 2021)
14. Saha, P., Mishra, D., Hernandez-Cruz, N., et al.: Self-supervised normality learning and divergence vector-guided model merging for zero-shot congenital heart disease detection in fetal ultrasound videos. In: Gee, J.C., et al. (eds.) MICCAI 2025. pp. 560–571. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-04981-0_53
15. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: Wang, Q., et al. (eds.) Machine Learning in Medical Imaging. pp. 379–387. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_44
16. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
17. Song, Y., Du, X., Zhang, Y., et al.: Multi-shot prototype contrastive learning and semantic reasoning for medical image segmentation. In: Greenspan, H., et al. (eds.) MICCAI 2023. pp. 578–588. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-43901-8_55
18. Stanford AIMI: Thyroid ultrasound cine-clip. https://doi.org/10.71718/7m5n-rh16
19. Tang, S., Yan, S., Qi, X., et al.: Few-shot medical image segmentation with high-fidelity prototypes. Medical Image Analysis **100**, 103412 (2025). https://doi.org/10.1016/j.media.2024.103412
20. Tversky, A.: Features of similarity. Psychological review **84**(4), 327 (1977). https://doi.org/10.1037/0033-295X.84.4.327
21. Tversky, A., Gati, I.: Similarity, separability, and the triangle inequality. Psychological review **89**(2), 123 (1982). https://doi.org/10.1037/0033-295X.89.2.123
22. Vega, D., Ceballos, H., Rincon, J.S.V., et al.: CARDIUM: Congenital anomaly recognition with diagnostic images and unified medical records. In: Third Workshop on Computer Vision for Automated Medical Diagnosis - arXiv:2510.15208 (2025)
23. Wang, J., Fan, L., Jing, W., et al.: Hypergraph tversky-aware domain incremental learning for brain tumor segmentation with missing modalities. In: Gee, J.C., et al. (eds.) MICCAI 2025. pp. 283–293. Springer Nature Switzerland (2026). https://doi.org/10.1007/978-3-032-05141-7_28
24. Yamashita, R., Kapoor, T., Alam, M.N., et al.: Toward reduction in false-positive thyroid nodule biopsies with a deep learning–based risk stratification system using us cine-clip images. Radiology: Artificial Intelligence **4**(3) (2022). https://doi.org/10.1148/ryai.210174

25. Zhang, S., Xu, Y., Usuyama, N., et al.: A multimodal biomedical foundation model trained from fifteen million image–text pairs. NEJM AI **2**(1), AIoa2400640 (2025). https://doi.org/10.1056/AIoa2400640