FROM FITTING TO FLOWING: RELIABILITY-GUIDED GEOMETRIC CORRECTION DYNAMICS FOR MULTI-VIEW PARTIAL MULTI-LABEL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view partial multi-label learning (MVPML) learns from heterogeneous views and ambiguous candidate label sets containing both relevant and irrelevant labels. Existing methods typically refine candidate supervision or jointly optimize disambiguation and classification, but ultimately treat inferred supervision as a target to fit. Errors in this supervision can therefore propagate directly into prediction, while multi-view geometry is mainly used for refinement or regularization rather than for controlling such propagation. We propose GeoCURVE, a reliability-guided, geometry-embedded continuous-time dynamical framework that instead treats inferred supervision as a controlled correction to a structural reference. Specifically, prediction evolves according to a reference-restored ordinary differential equation (ODE). Candidate reliability, indicating the likelihood that a candidate label is truly relevant, determines the correction signal, shared multi-view geometry governs how it propagates across structurally related instances, and a restoration term controls how much of it is retained. The resulting dynamics admit a unique globally exponentially stable equilibrium with a closed-form solution, exposing how reliability, geometry, and restoration jointly determine the final prediction. We derive an oracle-based error bound quantifying controlled uncertainty propagation and an incremental generalization bound showing how spectral geometry and restoration control correction complexity and the correction-induced generalization gap. Experiments on six benchmark datasets under multiple ambiguity settings demonstrate consistent improvements over representative MVPML methods, achieving relative AP improvements of up to 13% over competitive baselines on several datasets, while ablations show that controlling correction extent, rather than merely refining supervision, drives the gain.

1 INTRODUCTION

Multi-view partial multi-label learning (Chen et al., 2020; Yuan et al., 2023; Wang et al., 2025; Dong et al., 2027) considers a weakly supervised setting (Zhou, 2018) where each instance has multiple heterogeneous views and an ambiguous candidate label set containing all relevant labels together with irrelevant ones. Since exhaustive multi-label annotation is costly, such candidate annotations provide a practical form of weak supervision.

Existing MVPML methods mainly follow two paradigms: two-stage methods refine candidate supervision before classifier learning (Chen et al., 2020; Wu et al., 2020; Wang & Xu, 2024), while joint-learning methods couple supervision disambiguation with classifier optimization (Xu et al., 2023; Wang et al., 2025; Dong et al., 2027; Wu et al., 2025). Despite this difference, both ultimately treat inferred supervision as a target to fit. Because such supervision remains uncertain, its errors can propagate to the final prediction. The key challenge is therefore not only how to refine candidate supervision, but also how to control the influence of its residual uncertainty on prediction.

Since MVPML ultimately aims to estimate prediction scores under ambiguous supervision, we treat these scores as the modeling object rather than merely the output of learning, with the goal of controlling their uncertainty.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

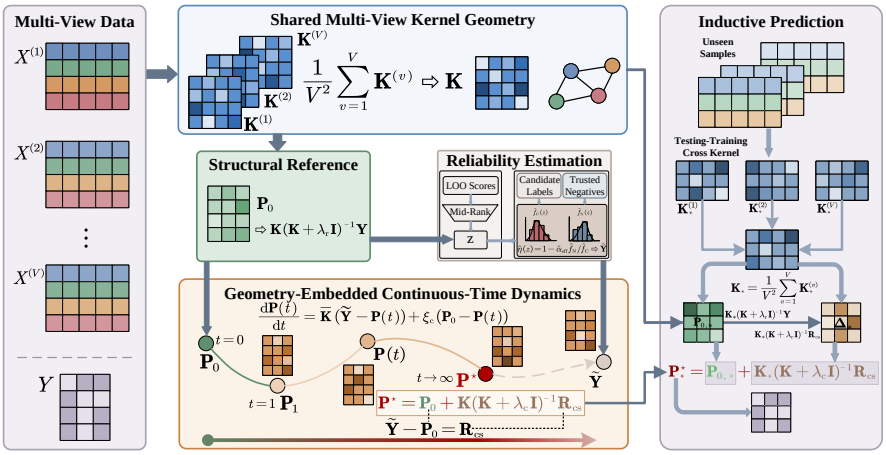


Figure 1: Overview of GeoCURVE. Candidate reliability defines the correction signal, shared multi-view geometry regulates its evolution, and reference restoration controls the final equilibrium.

Multi-view geometry provides structural information through inter-instance relations across heterogeneous views, and has been widely used for candidate-label refinement, label propagation, and structural regularization (Wu et al., 2020; Chen et al., 2020; Wang & Xu, 2024; Xu et al., 2023; Yuan et al., 2023; Wu et al., 2025; Dong et al., 2027). However, it is mainly used to improve supervision estimation or enforce consistency, leaving an important question underexplored: can geometry control how uncertain supervision influences prediction? We address this question by treating inferred supervision as a geometry-regulated correction to a structural reference, rather than as a target to fit.

Based on this principle, we propose GeoCURVE, as shown in Fig. 1, a reliability-guided, geometry-embedded dynamical framework for MVPML. GeoCURVE models prediction as a continuous-time state evolving from a structural reference: candidate reliability determines the correction signal, multi-view geometry governs how the correction propagates across structurally related instances, and a restoration term controls how much correction is retained. These components are unified in a reference-restored ordinary differential equation (ODE), whose unique globally stable equilibrium admits a closed-form solution.

Specifically, GeoCURVE constructs a candidate-specific correction from estimated label reliability and integrates view-specific kernels into a shared geometry embedded in the dynamical operator. The induced spectral regulation controls how correction components contribute to prediction, while restoration limits their overall influence. We derive an oracle-based error bound on the propagation of supervision uncertainty and an incremental generalization bound showing how spectral geometry and restoration regulate correction complexity and the correction-induced generalization gap. Experiments on six benchmark datasets under multiple ambiguity settings demonstrate consistent improvements over representative MVPML methods, while ablations show that controlling correction extent is more consistently beneficial than merely refining supervision. In summary, our main contributions are as follows:

- **Controlled-correction principle.** We treat inferred supervision as a controlled correction to a structural reference rather than a target to fit, explicitly regulating the influence of supervision uncertainty.
- **GeoCURVE framework.** We develop a reference-restored ODE integrating candidate reliability, shared multi-view geometry, and restoration, with a unique globally exponentially stable closed-form equilibrium.
- **Theoretical guarantees.** We derive an oracle error bound for supervision-uncertainty propagation and an incremental generalization bound showing how spectral geometry and restoration regulate correction complexity and the correction-induced generalization gap.
- **Empirical validation.** Experiments across six benchmarks and multiple ambiguity settings demonstrate consistent gains, while ablations highlight the importance of correction control beyond supervision refinement.

2 Related Work

Existing MVPML methods mainly resolve candidate-label ambiguity by exploiting structural information across multiple views. Early two-stage approaches, such as GRADIS (Chen et al., 2020), FIMAN (Wu et al., 2020), and TFMDD (Wang & Xu, 2024), first use multi-view graph or manifold structures to refine candidate supervision and then learn predictors from the refined labels. More recent methods integrate multi-view structure and supervision recovery more tightly within unified learning frameworks, including GMPM (Yuan et al., 2023), GLADE (Xu et al., 2023), VADIS (Wang et al., 2025), ICGC (Wu et al., 2025), and TARGET (Dong et al., 2027). Despite different formulations, these methods remain centered on estimating or optimizing latent supervision that directly participates in prediction learning. GeoCURVE instead treats inferred candidate-label evidence as a correction to a structural reference, with multi-view geometry regulating its evolution and restoration controlling its long-term influence. Continuous-time graph learning has been studied through graph neural differential equations and diffusion processes (Xhonneux et al., 2020; Chamberlain et al., 2021; Liu et al., 2025; Qin et al., 2026). These methods model continuous representation propagation over graph structures, whereas GeoCURVE introduces continuous-time dynamics into weakly supervised prediction to control how uncertain candidate-label evidence influences prediction.

3 Method

Consider n training instances represented by V views over C labels, where $\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}$ denotes the v-th feature matrix. Let $\mathbf{Y} \in \{0,1\}^{n\times C}$ and $\mathbf{Y}^{\mathrm{true}} \in \{0,1\}^{n\times C}$ denote the observed candidate-label and latent ground-truth matrices, respectively. Partial multi-label supervision assumes $\mathbf{Y}^{\mathrm{true}} \leq \mathbf{Y}$ element-wise, so $Y_{ij} = 0$ is a trusted negative whereas $Y_{ij} = 1$ denotes an ambiguous candidate positive. Given $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$ and $\mathbf{Y}$, our goal is to learn an inductive predictor whose training scores are collected in $\mathbf{P} \in \mathbb{R}^{n\times C}$, with larger values indicating stronger relevance.

GeoCURVE treats inferred supervision as a controlled correction to a structural reference rather than a target to fit. It first constructs a shared multi-view geometry and uses it to obtain a capacity-controlled structural reference $\mathbf{P}_0$. Estimated candidate reliability reweights the ambiguous supervision to form reliability-informed supervision $\widetilde{\mathbf{Y}}$, whose deviation from $\mathbf{P}_0$ defines the candidate-specific correction $\mathbf{R}_{\mathrm{cs}} = \widetilde{\mathbf{Y}} - \mathbf{P}_0$. GeoCURVE then formulates prediction as a continuous-time evolution from $\mathbf{P}_0$, driven by $\mathbf{R}_{\mathrm{cs}}$ and regulated by the multi-view geometry, while reference restoration controls correction retention at equilibrium. The final reference-plus-correction prediction extends to unseen samples via the shared test–training cross-kernel.

3.1 Shared Multi-View Kernel Geometry

GeoCURVE integrates inter-instance relations from different views into a shared geometry used for reference construction, prediction evolution, and inductive extension. For the v-th view, we construct a Gaussian radial basis function (RBF) kernel (Schölkopf & Smola, 2002): $K_{ij}^{(v)} = \exp(-\|\mathbf{x}_i^{(v)} - \mathbf{x}_j^{(v)}\|_2^2/(2\sigma_v^2))$, and fuse them as $\mathbf{K} = \frac{1}{V^2}\sum_{v=1}^{V}\mathbf{K}^{(v)}$. Since each view kernel is positive semidefinite, $\mathbf{K}$ is also a valid kernel.

3.2 Capacity-Controlled Structural Reference

Before estimating candidate-specific correction, GeoCURVE constructs a structural reference from which prediction evolves and relative to which correction is defined. Since the supervision $\mathbf{Y}$ contains ambiguous candidate positives, we obtain this reference by kernel ridge regression (KRR) (Fan et al., 2020) over the shared geometry. With kernel coefficient matrix $\mathbf{A}$, the KRR objective is

$$\mathbf{A}^{\star} = \arg\min_{\mathbf{A}} \|\mathbf{K}\mathbf{A} - \mathbf{Y}\|_F^2 + \lambda_r \operatorname{tr}(\mathbf{A}^T\mathbf{K}\mathbf{A}). \tag{1}$$

The corresponding fitted reference is

$$\mathbf{P}_0 = \mathbf{K}\mathbf{A}^{\star} = \mathbf{H}_r\mathbf{Y}, \qquad \mathbf{H}_r = \mathbf{K}(\mathbf{K} + \lambda_r\mathbf{I})^{-1}, \tag{2}$$

where $\mathbf{H}_r$ is the reference filter and $\lambda_r > 0$ controls its regularization. Thus, $\mathbf{P}_0$ is a geometry-smoothed reference rather than a label target; correction is defined relative to it.

Instead of tuning λ_r directly, GeoCURVE parameterizes reference flexibility by its normalized effective capacity. Let $\{\mu_i\}_{i=1}^n$ be the eigenvalues of $\mathbf{K}$. Since $\mathbf{H}_r$ has spectral response $\mu_i/(\mu_i+\lambda_r)$, we define $\gamma_r(\lambda_r) = \frac{1}{n}\operatorname{tr}(\mathbf{H}_r) = \frac{1}{n}\sum_{i=1}^n \frac{\mu_i}{\mu_i+\lambda_r}$. Thus, γ_r measures the average spectral response retained by the reference, with larger values corresponding to greater flexibility. Since $\gamma_r(\lambda_r)$ is continuous and strictly decreasing, each feasible γ_r uniquely determines λ_r (Appendix B.2).

3.3 Reliability-Guided Candidate-Specific Correction

The structural reference provides the starting point of prediction evolution, but candidate positives may differ substantially in reliability. GeoCURVE therefore estimates candidate reliability from structural evidence via a trusted-negative density-ratio model and defines correction relative to $\mathbf{P}_0$.

Since each in-sample reference score depends partly on its own labels, we remove this self-influence using the exact leave-one-out (LOO) prediction of the linear smoother (Golub et al., 1979):

$$\mathbf{P}_{0,i}^{(-i)} = (\mathbf{P}_{0,i} - H_{r,ii}\mathbf{Y}_i)/(1 - H_{r,ii}). \tag{3}$$

We then convert all nC LOO scores $\mathbf{P}_{0,i}^{(-i)}$ into comparable structural evidence by the pooled mid-rank transformation $z_{ij} = [\mathrm{midrank}(P_{0,ij}^{(-i)}) - \frac{1}{2}]/(nC) \in (0,1)$.

Trusted negatives provide a reference for negative-like evidence, whereas candidate entries contain both false and relevant labels. Let $f_N(z)$ and $f_C(z)$ denote the corresponding densities. Motivated by the observation that false candidates often share structural characteristics with negative labels, we model candidate evidence using a negative-like mixture assumption.

$$f_C(z) \approx \alpha_{\text{eff}} f_N(z) + (1 - \alpha_{\text{eff}}) f_T(z), \tag{4}$$

where $f_T(z)$ is the latent evidence density of relevant candidates and $\alpha_{\text{eff}} \in [0,1]$ is the effective negative-like proportion. This assumption is empirically assessed in Sec. 5.5.

Let G_C and G_N denote the corresponding cumulative distributions. In the low-evidence region, where the relevant-candidate contribution becomes negligible, $G_C(t)/G_N(t)$ approaches α_{eff} while remaining no smaller than it for $G_N(t) > 0$ (see Appendix B.4). We therefore estimate

$$\widehat{\alpha}_{\text{eff}} = \mathrm{clip}\big(\min_{t\in\mathcal{T}} \widehat{G}_C(t)/\widehat{G}_N(t),\, 0,\, 1\big). \tag{5}$$

The ratio $\widehat{\alpha}_{\text{eff}}\widehat{f}_N(z)/\widehat{f}_C(z)$ then estimates the local negative-like contribution within candidate evidence, giving reliability

$$\widehat{\eta}(z) = \mathrm{clip}\big(1 - \widehat{\alpha}_{\text{eff}}\widehat{f}_N(z)/\widehat{f}_C(z),\, 0,\, 1\big). \tag{6}$$

Finally, we use the estimated reliability to reweight candidate supervision:

$$\widetilde{Y}_{ij} = \begin{cases} \widehat{\eta}(z_{ij}), & Y_{ij} = 1, \\ 0, & Y_{ij} = 0, \end{cases} \qquad \mathbf{R}_{\text{cs}} = \widetilde{\mathbf{Y}} - \mathbf{P}_0. \tag{7}$$

Thus, $\widetilde{\mathbf{Y}}$ is reliability-informed supervision, while $\mathbf{R}_{\text{cs}}$ is the candidate-specific correction relative to the structural reference.

3.4 Geometry-Embedded Continuous-Time Dynamics

We next formulate how candidate-specific correction drives prediction evolution. Rather than directly defining a static corrected estimator, GeoCURVE characterizes correction as a continuous-time process whose equilibrium determines the final prediction. This dynamical formulation reveals how candidate uncertainty is progressively propagated by the shared geometry and how reference restoration regulates the retained correction extent.

We first analyze geometry-only dynamics and show that, although geometry regulates the evolution, long-term dynamics eventually absorb the entire correction along every non-null kernel direction, including residual uncertainty inherited from inferred supervision. This motivates reference restoration, leading to the final GeoCURVE dynamics and a closed-form equilibrium with controlled correction extent.

Geometry-Only Corrective Dynamics. The candidate-specific correction $\mathbf{R}_{\mathrm{cs}}$ indicates how prediction should move away from the structural reference, but not how this correction should evolve across samples. GeoCURVE therefore embeds the shared geometry into the dynamics using the normalized kernel $\overline{\mathbf{K}} = \mathbf{K}/\mu_{\max}$, where $\mu_{\max}$ is the largest eigenvalue of $\mathbf{K}$.

Define the corrective energy as $\mathcal{E}_0(\mathbf{P}) = \frac{1}{2}\|\mathbf{P} - \widetilde{\mathbf{Y}}\|_F^2$. Propagating its negative gradient through the shared geometry yields

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = -\overline{\mathbf{K}}\nabla_{\mathbf{P}}\mathcal{E}_0(\mathbf{P}(t)) = \overline{\mathbf{K}}\big(\widetilde{\mathbf{Y}} - \mathbf{P}(t)\big), \qquad \mathbf{P}(0) = \mathbf{P}_0. \tag{8}$$

Since $\overline{\mathbf{K}} \succeq 0$, $\mathcal{E}_0$ is non-increasing along the trajectory. Moreover, $\widetilde{\mathbf{Y}} - \mathbf{P}(0) = \mathbf{R}_{\mathrm{cs}}$, so the correction determines the initial evolution direction, while $\overline{\mathbf{K}}$ regulates how it evolves across samples.

The resulting trajectory has the closed form

$$\mathbf{P}(t) = \mathbf{P}_0 + \big(\mathbf{I} - e^{-t\overline{\mathbf{K}}}\big)\mathbf{R}_{\mathrm{cs}}. \tag{9}$$

The matrix filter $\mathbf{I} - e^{-t\overline{\mathbf{K}}}$ has spectral response $1 - e^{-t\nu_s}$ along the s-th eigendirection, where $\nu_s = \mu_s/\mu_{\max} \in [0, 1]$ is the corresponding eigenvalue of $\overline{\mathbf{K}}$. Hence, larger ν_s yields faster correction in that direction, whereas $\nu_s = 0$ produces no correction there.

However, for every $\nu_s > 0$, $1 - e^{-t\nu_s} \to 1$ as $t \to \infty$. Thus, long-term evolution fully incorporates every correction component in the non-null kernel subspace; if $\mathbf{K}$ is full rank, $\mathbf{P}(t) \to \widetilde{\mathbf{Y}}$. Geometry therefore regulates how correction evolves but cannot limit its long-term influence. Since $\widetilde{\mathbf{Y}}$ is inferred from ambiguous supervision, residual uncertainty may also enter the limiting prediction, motivating reference restoration.

Reference-Restored Dynamics and Equilibrium. The geometry-only dynamics eventually incorporate every correction component in the non-null kernel subspace. To limit this long-term influence, GeoCURVE introduces reference restoration:

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = \overline{\mathbf{K}}\big(\widetilde{\mathbf{Y}} - \mathbf{P}(t)\big) + \xi_c\big(\mathbf{P}_0 - \mathbf{P}(t)\big), \qquad \mathbf{P}(0) = \mathbf{P}_0, \tag{10}$$

where $\xi_c > 0$ controls the restoration strength. The first term drives prediction toward the reliability-informed supervision through the shared geometry, whereas the second pulls it toward the structural reference. Together, they balance correction and reference preservation, limiting the influence of residual supervision uncertainty. At initialization, the restoration term vanishes and $\mathrm{d}\mathbf{P}(0)/\mathrm{d}t = \overline{\mathbf{K}}\mathbf{R}_{\mathrm{cs}}$, so restoration preserves the initial correction direction while regulating subsequent evolution.

Let $\lambda_c = \xi_c\mu_{\max}$. Using $\widetilde{\mathbf{Y}} = \mathbf{P}_0 + \mathbf{R}_{\mathrm{cs}}$, the dynamics admit the closed-form trajectory

$$\mathbf{P}(t) = \mathbf{P}_0 + \Big[\mathbf{I} - e^{-t(\overline{\mathbf{K}} + \xi_c\mathbf{I})}\Big]\mathbf{H}_c\mathbf{R}_{\mathrm{cs}}, \qquad \mathbf{H}_c = \mathbf{K}(\mathbf{K} + \lambda_c\mathbf{I})^{-1}. \tag{11}$$

Here, $\mathbf{H}_c$ filters the correction, while the matrix exponential governs its temporal incorporation.

Since $\overline{\mathbf{K}} + \xi_c\mathbf{I} \succ 0$, $e^{-t(\overline{\mathbf{K}} + \xi_c\mathbf{I})} \to \mathbf{0}$ as $t \to \infty$, so the dynamics converge to the unique equilibrium

$$\mathbf{P}^\star = \lim_{t\to\infty}\mathbf{P}(t) = \mathbf{P}_0 + \mathbf{H}_c\mathbf{R}_{\mathrm{cs}}. \tag{12}$$

Along a kernel eigendirection with eigenvalue μ_s, $\mathbf{H}_c$ has spectral response $g_s = \mu_s/(\mu_s + \lambda_c)$. Hence, larger μ_s retains more correction in that direction, whereas larger λ_c shifts the equilibrium toward the structural reference (see Appendix B.6). The normalized correction capacity $\gamma_c(\lambda_c) = \frac{1}{n}\operatorname{tr}(\mathbf{H}_c) = \frac{1}{n}\sum_{s=1}^{n}\mu_s/(\mu_s + \lambda_c)$ summarizes the overall correction extent.

3.5 Inductive Prediction

For unseen samples, GeoCURVE extends the learned reference and correction through their similarities to training instances. Let $\mathbf{K}_*^{(v)}$ denote the test–training cross-kernel in the v-th view. Following the training fusion rule, we form $\mathbf{K}_* = \frac{1}{V^2}\sum_{v=1}^{V}\mathbf{K}_*^{(v)}$, which preserves the multi-view geometry.

Using the kernel ridge extension (Saunders et al., 1998), we obtain

$$\mathbf{P}_{0,*} = \mathbf{K}_*(\mathbf{K} + \lambda_r \mathbf{I})^{-1}\mathbf{Y}, \qquad \mathbf{\Delta}_* = \mathbf{K}_*(\mathbf{K} + \lambda_c \mathbf{I})^{-1}\mathbf{R}_{\mathrm{cs}}, \qquad \mathbf{P}^\star_* = \mathbf{P}_{0,*} + \mathbf{\Delta}_*. \tag{13}$$

Thus, GeoCURVE extends both the structural reference and correction to unseen samples through the shared geometry. Rather than directly fitting inferred supervision, GeoCURVE models prediction scores as dynamical states and controls uncertainty propagation through continuous evolution. As illustrated in Fig. 1, GeoCURVE constructs shared geometry, derives the reference and correction, evolves prediction through reference-restored dynamics, and performs inductive prediction via cross-kernel evaluation. The complete procedure is summarized in Algorithm 1 in Appendix A.

4 Theoretical Analysis

In this section, we analyze GeoCURVE from dynamical and statistical perspectives, showing how supervision uncertainty is controlled and geometry regulates correction complexity.

Theorem 1. *Fix the shared kernel $\mathbf{K}$ and regularization parameters λ_r, λ_c. Let $\mathbf{P}_0^{\mathrm{or}} = \mathbf{H}_r\mathbf{Y}^{\mathrm{true}}$ denote the oracle structural reference, and define the corresponding oracle equilibrium as $\mathbf{P}^\star_{\mathrm{or}} = \mathbf{P}_0^{\mathrm{or}} + \mathbf{H}_c(\mathbf{Y}^{\mathrm{true}} - \mathbf{P}_0^{\mathrm{or}})$. Then*

$$\|\mathbf{P}^\star - \mathbf{P}^\star_{\mathrm{or}}\|_F \le \rho_{r,c}\|\mathbf{Y} - \mathbf{Y}^{\mathrm{true}}\|_F + \vartheta_c\|\widetilde{\mathbf{Y}} - \mathbf{Y}^{\mathrm{true}}\|_F, \tag{14}$$

where $\rho_{r,c} = \max_s \lambda_c\mu_s/[(\mu_s+\lambda_c)(\mu_s+\lambda_r)] < 1$ and $\vartheta_c = \mu_{\max}/(\mu_{\max}+\lambda_c) = 1/(1+\xi_c) < 1$. Moreover,

$$\|\mathbf{P}^\star - \mathbf{Y}^{\mathrm{true}}\|_F \le \|\mathbf{P}^\star_{\mathrm{or}} - \mathbf{Y}^{\mathrm{true}}\|_F + \rho_{r,c}\|\mathbf{Y} - \mathbf{Y}^{\mathrm{true}}\|_F + \vartheta_c\|\widetilde{\mathbf{Y}} - \mathbf{Y}^{\mathrm{true}}\|_F. \tag{15}$$

Theorem 1 separates the effect of ambiguous supervision into reference contamination and inferred-supervision error. Since both $\rho_{r,c}$ and ϑ_c are strictly smaller than one, neither uncertainty channel is amplified in the final equilibrium; the prediction error is bounded by the oracle error plus contracted penalties from the two sources of supervision uncertainty. This result provides theoretical support for GeoCURVE's ability to control uncertainty from ambiguous supervision (Proof in Appendix C.1).

For the squared loss with respect to the latent true labels, define

$$\mathcal{R}(\mathbf{f}) = \mathbb{E}[\frac{1}{C}\|\mathbf{f}(X) - \mathbf{Y}^{\mathrm{true}}\|_2^2], \qquad \widehat{\mathcal{R}}(\mathbf{f}) = \frac{1}{nC}\sum_{i=1}^n \|\mathbf{f}(X_i) - \mathbf{Y}_i^{\mathrm{true}}\|_2^2. \tag{16}$$

We further define the corresponding risk increments as $\Delta\mathcal{R}(\mathbf{h}) = \mathcal{R}(\mathbf{f}_0+\mathbf{h}) - \mathcal{R}(\mathbf{f}_0)$ and $\Delta\widehat{\mathcal{R}}(\mathbf{h}) = \widehat{\mathcal{R}}(\mathbf{f}_0 + \mathbf{h}) - \widehat{\mathcal{R}}(\mathbf{f}_0)$.

Theorem 2. *Assume that $\{(X_i, \mathbf{Y}_i^{\mathrm{true}})\}_{i=1}^n$ are i.i.d. with $\mathbf{Y}_i^{\mathrm{true}} \in \{0,1\}^C$, and that the shared kernel k, including its preprocessing and bandwidths, is fixed independently of the sample and satisfies $k(x,x) \le \kappa^2$. Let $\mathbf{f}^\star = \mathbf{f}_0 + \mathbf{h}_c$. For $\mathbf{K} = \mathbf{U}\,\mathrm{diag}(\mu_1, \dots, \mu_n)\mathbf{U}^T$, where $\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s} = \mathbf{u}_s^T\mathbf{R}_{\mathrm{cs}}$ denotes the correction component along the s-th eigendirection,*

$$\|\mathbf{h}_c\|^2_{\mathcal{H}_k^C} = \sum_{\mu_s>0} \frac{\mu_s}{(\mu_s+\lambda_c)^2}\|\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s}\|_2^2. \tag{17}$$

Moreover, for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$|\Delta\mathcal{R}(\mathbf{h}_c) - \Delta\widehat{\mathcal{R}}(\mathbf{h}_c)| \le \Psi_n(q_0, q_c, \delta), \tag{18}$$

where $q_0 = \|\mathbf{f}_0\|_{\mathcal{H}_k^C}$, $q_c = \|\mathbf{h}_c\|_{\mathcal{H}_k^C}$, and

$$\Psi_n(q_0, q_c, \delta) = \frac{1 + \sqrt{2\log(2/\delta)}}{C\sqrt{n}}\Big[\kappa^2(2q_0q_c + q_c^2) + 2\kappa\sqrt{C}\,q_c\Big]. \tag{19}$$

Furthermore, $\Psi_n(q_0, q_c, \delta)$ is nondecreasing in q_c.

Eq. (17) shows that the shared geometry controls correction complexity through both the spectral decomposition $\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs}} = \mathbf{U}^T\mathbf{R}_{\mathrm{cs}}$ and the weights $\mu_s/(\mu_s+\lambda_c)^2$. For fixed λ_c, the spectral weights

are maximized at $\mu_s = \lambda_c$: they increase when $\mu_s < \lambda_c$ and decrease when $\mu_s > \lambda_c$. Moreover, increasing λ_c decreases every nonzero spectral weight and yields $q_c^2 \leq \|\mathbf{R}_{\mathrm{cs}}\|_F^2/(4\lambda_c)$. Since $\Psi_n(q_0, q_c, \delta)$ is nondecreasing in q_c and decays as $O(n^{-1/2})$, stronger restoration or smaller correction energy yields a tighter incremental generalization bound, while the term $2q_0q_c$ further shows that the statistical cost of correction also depends on the complexity of the structural reference. Thus, GeoCURVE controls the influence of uncertain supervision through geometry-dependent spectral regulation and restoration-controlled correction complexity (Proof in Appendix C.2).

5 Experiments

5.1 Experimental Setup

Datasets and Compared Methods. We evaluate GeoCURVE on six multi-view multi-label benchmark datasets: Emotions (Trohidis et al., 2008), Yeast (Elisseeff & Weston, 2001), EspGame5k (von Ahn & Dabbish, 2004), OBJECT (Zhang et al., 2020), Human, and Plant (Xu, 2013). Following standard PML protocols (Wu et al., 2020; Xu et al., 2023; Cour et al., 2011; Liu & Dietterich, 2012; Yu & Zhang, 2016), we corrupt a proportion $p \in \{0.3, 0.7\}$ of instances by adding $r \in \{1, 2, 3\}$ irrelevant labels without removing relevant ones, yielding six ambiguity configurations per dataset. We compare GeoCURVE with five representative MVPML methods: GRADIS (Chen et al., 2020), FIMAN (Wu et al., 2020), GLADE (Xu et al., 2023), MSPML-TLR (Deng et al., 2024), and VADIS (Wang et al., 2025), together with TMvML (Zhong et al., 2025) and FastGRAIL (Liu et al., 2026) from the related multi-view multi-label and partial multi-label settings, respectively. Dataset statistics and baseline settings are provided in Appendix D.1 and Appendix D.2.

Implementation Details. We evaluate all methods using average precision (AP), Hamming loss (HL), one-error (OE), ranking loss (RL), and coverage (COV) (Zhang & Zhou, 2014). Ranking metrics use the predicted scores, while HL uses binary predictions obtained by thresholding at 0.75. Results are reported as mean $\pm$ standard deviation over the ten outer test folds. For GeoCURVE, γ_r and γ_c are jointly selected by nested cross-validation (Varma & Simon, 2006; Cawley & Talbot, 2010) using only weak supervision within each outer-training set, with SoftAP as the inner selection criterion. We search $\gamma_r \in \{0.05, 0.075, 0.10, 0.125, 0.15, 0.175, 0.20\}$ and $\gamma_c \in \{0.05, 0.075, 0.10, 0.125, 0.15, 0.175, 0.20, 0.225, 0.25, 0.275, 0.30, 0.325, 0.35, 0.375, 0.40\}$. The HL threshold is fixed at 0.75. All experiments are conducted with a fixed random seed of 42 for reproducibility. Feature preprocessing, kernel construction, and further implementation details are provided in Appendix D.3.

5.2 Comparison Results

Table 1 reports the AP results under different ambiguity configurations. GeoCURVE achieves the best AP in 33 of the 36 dataset–configuration cases and ranks first or second in all cases. It also ranks among the top two in 35, 36, 36, and 31 cases on OE, RL, COV, and HL, respectively; complete results are provided in Appendix D.4. These results demonstrate the effectiveness and consistent performance of GeoCURVE across different ambiguity configurations and evaluation criteria.

The Friedman test with the Iman–Davenport correction rejects equal performance for all five metrics at $\alpha = 0.05$ (Appendix D.5). The subsequent Bonferroni–Dunn test (Demšar, 2006) shows that GeoCURVE achieves the best average rank on AP, OE, RL, and COV and the second-best rank on HL (Appendix D.5). It significantly outperforms all competitors on RL and COV, is statistically comparable only to GRADIS on AP and OE, and to GRADIS and FastGRAIL on HL. These results further substantiate the overall effectiveness of GeoCURVE across different evaluation criteria.

5.3 Parameter Sensitivity

GeoCURVE exhibits smooth sensitivity to both capacity parameters γ_r and γ_c across different ambiguity configurations, without a sharply isolated optimum. Small-to-moderate capacity values are generally favorable, whereas overly large values may gradually reduce AP, particularly under stronger ambiguity; complete sensitivity results on all six datasets are provided in Appendix D.6.

Table 1: Average precision (AP ↑) results under different ambiguity configurations. The best and second-best results in each configuration are shown in bold and underlined, respectively.

Method	(p, r)	Emotions	EspGame5k	Yeast	OBJECT	Human	Plant
FastGRAIL	$(0.3, 1)$	0.785±0.028	0.389±0.008	0.754±0.013	0.703±0.012	0.589±0.012	0.571±0.023
	$(0.3, 2)$	0.781±0.027	0.387±0.010	0.754±0.013	0.701±0.012	0.589±0.012	0.565±0.023
	$(0.3, 3)$	0.781±0.024	0.383±0.011	0.752±0.013	0.701±0.012	0.586±0.017	0.556±0.026
	$(0.7, 1)$	0.774±0.023	0.384±0.008	0.753±0.012	0.702±0.012	0.590±0.014	0.557±0.014
	$(0.7, 2)$	0.782±0.024	0.376±0.008	0.753±0.013	0.699±0.012	0.586±0.018	0.553±0.021
	$(0.7, 3)$	0.765±0.031	0.371±0.011	0.749±0.013	0.694±0.011	0.578±0.015	0.533±0.014
TMvML	$(0.3, 1)$	0.773±0.028	0.397±0.012	0.745±0.013	0.640±0.012	0.607±0.012	0.578±0.016
	$(0.3, 2)$	0.764±0.026	0.395±0.011	0.742±0.014	0.635±0.009	0.602±0.010	0.573±0.017
	$(0.3, 3)$	0.765±0.032	0.396±0.012	0.742±0.014	0.634±0.013	0.606±0.010	0.564±0.026
	$(0.7, 1)$	0.776±0.021	0.394±0.010	0.743±0.013	0.635±0.011	0.600±0.011	0.569±0.029
	$(0.7, 2)$	0.760±0.025	0.398±0.012	0.741±0.015	0.631±0.014	0.593±0.013	0.547±0.027
	$(0.7, 3)$	0.747±0.022	0.403±0.006	0.736±0.014	0.630±0.014	0.587±0.011	0.542±0.019
GRADIS	$(0.3, 1)$	0.814±0.034	0.400±0.005	0.768±0.009	0.708±0.011	0.615±0.018	0.572±0.018
	$(0.3, 2)$	**0.818±0.029**	0.396±0.007	0.767±0.012	0.708±0.010	0.616±0.020	0.560±0.029
	$(0.3, 3)$	**0.813±0.028**	0.390±0.004	0.764±0.011	0.705±0.010	0.614±0.020	0.552±0.036
	$(0.7, 1)$	**0.818±0.037**	0.394±0.008	0.765±0.013	0.704±0.011	0.611±0.016	0.559±0.024
	$(0.7, 2)$	0.791±0.030	0.387±0.007	0.762±0.011	0.697±0.012	0.606±0.011	0.552±0.024
	$(0.7, 3)$	0.768±0.029	0.378±0.008	0.755±0.014	0.692±0.013	0.597±0.022	0.550±0.032
FIMAN	$(0.3, 1)$	0.792±0.029	0.397±0.008	0.756±0.013	0.695±0.014	0.586±0.013	0.492±0.020
	$(0.3, 2)$	0.792±0.028	0.393±0.006	0.754±0.013	0.693±0.015	0.583±0.014	0.482±0.037
	$(0.3, 3)$	0.792±0.024	0.389±0.007	0.754±0.014	0.692±0.014	0.575±0.016	0.469±0.024
	$(0.7, 1)$	0.786±0.032	0.391±0.006	0.753±0.013	0.691±0.012	0.573±0.012	0.463±0.025
	$(0.7, 2)$	0.775±0.018	0.380±0.007	0.754±0.013	0.687±0.014	0.552±0.014	0.447±0.023
	$(0.7, 3)$	0.772±0.027	0.371±0.009	0.752±0.013	0.685±0.012	0.545±0.020	0.429±0.030
GLADE	$(0.3, 1)$	0.804±0.033	0.391±0.007	0.760±0.014	0.699±0.015	0.577±0.015	0.468±0.029
	$(0.3, 2)$	0.792±0.032	0.385±0.006	0.757±0.014	0.696±0.016	0.570±0.016	0.457±0.041
	$(0.3, 3)$	0.788±0.029	0.379±0.005	0.756±0.014	0.692±0.015	0.562±0.024	0.438±0.034
	$(0.7, 1)$	0.776±0.035	0.383±0.008	0.757±0.014	0.693±0.014	0.565±0.016	0.440±0.033
	$(0.7, 2)$	0.765±0.036	0.370±0.008	0.752±0.015	0.686±0.014	0.533±0.018	0.409±0.022
	$(0.7, 3)$	0.732±0.030	0.357±0.011	0.746±0.014	0.677±0.014	0.512±0.016	0.381±0.044
MSPML-TLR	$(0.3, 1)$	0.758±0.042	0.273±0.009	0.694±0.014	0.143±0.002	0.513±0.016	0.490±0.033
	$(0.3, 2)$	0.740±0.034	0.267±0.008	0.691±0.014	0.143±0.002	0.507±0.020	0.483±0.027
	$(0.3, 3)$	0.732±0.028	0.268±0.008	0.687±0.014	0.143±0.002	0.496±0.015	0.472±0.031
	$(0.7, 1)$	0.733±0.037	0.269±0.009	0.692±0.012	0.143±0.002	0.499±0.017	0.490±0.026
	$(0.7, 2)$	0.731±0.036	0.260±0.008	0.682±0.009	0.142±0.005	0.463±0.014	0.466±0.030
	$(0.7, 3)$	0.636±0.040	0.245±0.006	0.674±0.014	0.212±0.139	0.420±0.014	0.444±0.029
VADIS	$(0.3, 1)$	0.803±0.033	0.323±0.013	0.761±0.014	0.700±0.014	0.582±0.013	0.489±0.022
	$(0.3, 2)$	0.788±0.028	0.321±0.013	0.759±0.014	0.697±0.013	0.576±0.018	0.475±0.042
	$(0.3, 3)$	0.788±0.026	0.320±0.012	0.758±0.014	0.696±0.012	0.567±0.023	0.455±0.036
	$(0.7, 1)$	0.781±0.039	0.320±0.012	0.759±0.013	0.698±0.013	0.570±0.018	0.457±0.025
	$(0.7, 2)$	0.767±0.032	0.317±0.013	0.754±0.015	0.694±0.012	0.541±0.016	0.434±0.023
	$(0.7, 3)$	0.737±0.030	0.315±0.013	0.751±0.014	0.687±0.013	0.522±0.013	0.399±0.037
GeoCURVE	$(0.3, 1)$	**0.819±0.032**	**0.453±0.006**	**0.780±0.014**	**0.742±0.013**	**0.643±0.017**	**0.595±0.019**
	$(0.3, 2)$	0.810±0.030	**0.449±0.007**	**0.779±0.014**	**0.738±0.012**	**0.640±0.016**	**0.595±0.017**
	$(0.3, 3)$	0.812±0.030	**0.448±0.006**	**0.779±0.011**	**0.737±0.012**	**0.637±0.019**	**0.589±0.026**
	$(0.7, 1)$	0.809±0.032	**0.450±0.006**	**0.775±0.014**	**0.738±0.014**	**0.638±0.015**	**0.590±0.020**
	$(0.7, 2)$	**0.795±0.034**	**0.444±0.006**	**0.775±0.014**	**0.735±0.011**	**0.630±0.016**	**0.580±0.018**
	$(0.7, 3)$	**0.774±0.034**	**0.441±0.007**	**0.768±0.014**	**0.727±0.014**	**0.625±0.017**	**0.571±0.025**

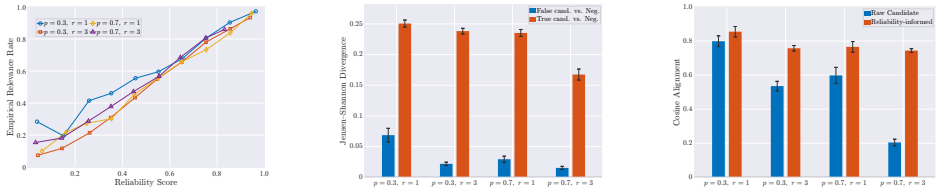
5.4 Ablation Study

To assess the contribution of each component, we construct five controlled variants: *w/o Correction* uses only the structural reference; *w/o LOO* uses in-sample reference predictions; *w/o Reliability* assigns equal candidate reliability; *w/o Geom.* replaces $\mathbf{K}$ with $\mathbf{I}$; and *w/o Restoration* removes the reference-restoring term. All variants reuse the capacities selected for the full model. Table 2 reports AP under four representative configurations covering low and high levels of both ambiguity factors.

Removing restoration causes the most consistent degradation, reducing AP in every displayed dataset–configuration pair and confirming the importance of limiting the long-term influence of uncertain correction. Removing correction also generally degrades performance, particularly on OBJECT, Yeast, and Human, showing that the structural reference alone cannot fully exploit weak supervision. The effects of LOO, reliability estimation, and geometry are more configuration-dependent, indicating that their benefits vary with ambiguity and dataset structure. Overall, the full model achieves the strongest aggregate performance across the displayed configurations, supporting the coordinated use of reliability-guided correction, multi-view geometry, and reference restoration.

Table 2: Ablation results of GeoCURVE in terms of Average precision (AP ↑) under four representative ambiguity configurations. The best result in each row is highlighted in bold.

Dataset	(p, r)	w/o Corr.	w/o LOO	w/o Reliab.	w/o Geom.	w/o Restor.	**GeoCURVE**
Emotions	$(0.3, 1)$	0.811±0.033	0.818±0.033	**0.821±0.032**	0.817±0.033	0.779±0.024	0.819±0.032
	$(0.3, 3)$	0.812±0.033	0.808±0.027	0.808±0.025	**0.817±0.028**	0.785±0.029	0.812±0.030
	$(0.7, 1)$	0.810±0.033	**0.811±0.032**	0.808±0.026	0.809±0.032	0.776±0.030	0.809±0.032
	$(0.7, 3)$	0.769±0.035	0.766±0.040	0.763±0.040	0.773±0.036	0.753±0.032	**0.774±0.034**
EspGame5k	$(0.3, 1)$	0.453±0.006	0.454±0.006	0.453±0.006	**0.456±0.007**	0.420±0.008	0.453±0.006
	$(0.3, 3)$	0.446±0.006	0.448±0.006	0.448±0.006	**0.450±0.006**	0.420±0.010	0.448±0.006
	$(0.7, 1)$	0.449±0.006	0.450±0.006	0.450±0.006	**0.453±0.007**	0.418±0.011	0.450±0.006
	$(0.7, 3)$	0.438±0.008	0.437±0.006	0.437±0.007	**0.441±0.008**	0.410±0.010	**0.441±0.007**
Yeast	$(0.3, 1)$	0.774±0.014	0.779±0.014	0.779±0.015	0.778±0.014	0.719±0.016	**0.780±0.014**
	$(0.3, 3)$	0.771±0.012	0.773±0.011	0.771±0.012	0.774±0.011	0.729±0.012	**0.779±0.011**
	$(0.7, 1)$	0.772±0.014	0.773±0.014	0.772±0.013	0.772±0.013	0.715±0.014	**0.775±0.014**
	$(0.7, 3)$	0.767±0.012	0.760±0.014	0.755±0.015	0.767±0.014	0.717±0.016	**0.768±0.014**
OBJECT	$(0.3, 1)$	0.727±0.012	**0.743±0.012**	0.741±0.013	0.737±0.011	0.694±0.010	0.742±0.013
	$(0.3, 3)$	0.725±0.012	**0.738±0.012**	0.736±0.012	0.736±0.012	0.693±0.011	0.737±0.012
	$(0.7, 1)$	0.725±0.013	0.737±0.013	0.736±0.013	0.735±0.013	0.692±0.011	**0.738±0.014**
	$(0.7, 3)$	0.724±0.015	0.722±0.015	0.722±0.014	**0.728±0.014**	0.692±0.013	0.727±0.014
Human	$(0.3, 1)$	0.638±0.012	0.641±0.017	0.640±0.018	0.640±0.015	0.606±0.020	**0.643±0.017**
	$(0.3, 3)$	0.636±0.015	0.632±0.024	0.632±0.024	**0.637±0.015**	0.603±0.024	**0.637±0.019**
	$(0.7, 1)$	0.637±0.018	0.635±0.019	0.635±0.020	**0.638±0.019**	0.603±0.019	**0.638±0.015**
	$(0.7, 3)$	0.624±0.018	0.610±0.025	0.609±0.024	0.623±0.019	0.594±0.020	**0.625±0.017**
Plant	$(0.3, 1)$	**0.598±0.021**	0.597±0.016	**0.598±0.015**	0.592±0.016	0.530±0.035	0.595±0.019
	$(0.3, 3)$	0.588±0.030	**0.594±0.025**	0.593±0.024	0.589±0.027	0.541±0.024	0.589±0.026
	$(0.7, 1)$	**0.590±0.029**	0.588±0.021	0.588±0.021	0.588±0.027	0.528±0.032	**0.590±0.020**
	$(0.7, 3)$	0.573±0.024	0.564±0.022	0.559±0.023	**0.576±0.023**	0.498±0.038	0.571±0.025



(a) (b) (c)

Figure 2: Candidate reliability and correction on Emotions: reliability–relevance relation, Jensen–Shannon divergence under the negative-like mixture assumption, and oracle correction alignment.

5.5 Analysis of Candidate Reliability and Correction

We analyze candidate reliability estimation and its effect on correction. Fig. 2 shows that (a) higher reliability scores correspond to higher empirical relevance rates; (b) false candidates exhibit smaller Jensen–Shannon divergence (Lin, 1991) from trusted negatives than true candidates, consistent with the negative-like assumption; and (c) reliability-informed correction achieves higher cosine alignment with the oracle correction than raw candidate correction, improving in 223 of 240 outer folds across six datasets and four ambiguity settings. These results support reliability-guided candidate-specific correction, which provides the direction later regulated by shared multi-view geometry.

6 Conclusion

We proposed GeoCURVE, a reliability-guided, geometry-embedded continuous-time framework for MVPML that treats inferred supervision as a controlled correction to a structural reference rather than a target to fit. Candidate reliability determines the correction, multi-view geometry regulates its evolution, and reference restoration controls its final influence, yielding a closed-form equilibrium. Theoretically, we establish an oracle error bound for uncertainty propagation and an incremental generalization bound for correction complexity, while experiments on six benchmarks and multiple ambiguity settings demonstrate consistent improvements supported by ablations and diagnostics. Overall, GeoCURVE provides a controlled-correction perspective for weak supervision, where uncertain candidate evidence guides prediction without being directly imposed as the prediction target.

Ethics statement

This work does not involve human subjects, private data, or any data collection procedure requiring ethical approval. All experiments are conducted on publicly available benchmark datasets. The proposed method is designed for improving learning from ambiguous supervision and does not introduce additional privacy, security, or fairness risks beyond those inherent in the underlying datasets.

Reproducibility statement

We make efforts to ensure the reproducibility of GeoCURVE from algorithmic, theoretical, and experimental perspectives. The main paper presents the complete formulation of the proposed framework, including the shared multi-view kernel geometry (Section 3.1), capacity-controlled structural reference (Section 3.2), reliability-guided candidate-specific correction (Section 3.3), geometry-embedded continuous-time dynamics (Section 3.4), and inductive prediction procedure (Section 3.5). The complete training and prediction procedure is summarized in Appendix A.

Theoretical assumptions, model properties, and complete proofs are provided in Appendices B–C, including the validity of the shared multi-view kernel, capacity parameterization, dynamical well-posedness, spectral characterization, uncertainty propagation bound, and generalization analysis. Detailed experimental protocols, including dataset descriptions, baseline configurations, preprocessing procedures, hyperparameter selection strategies, and evaluation settings, are provided in Appendix D. The source code and implementation resources will be released through a public repository after acceptance.

References

Gavin C. Cawley and Nicola L. C. Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010.

Ben Chamberlain, James Rowbottom, Maria I. Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1407–1418. PMLR, 2021.

Ze-Sen Chen, Xuan Wu, Qing-Guo Chen, Yao Hu, and Min-Ling Zhang. Multi-view partial multi-label learning with graph-based disambiguation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 3553–3560, 2020. doi: 10.1609/aaai.v34i04.5761.

Timothee Cour, Ben Sapp, and Ben Taskar. Learning from partial labels. *Journal of Machine Learning Research*, 12(42):1501–1536, 2011.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.

Tingquan Deng, Yiying Chen, Taoli Yang, Ge Yang, and Ming Yang. Multi-source partial multi-label learning via tensor decomposition and nonconvex regularization. *Information Fusion*, 112: 102583, 2024. doi: 10.1016/j.inffus.2024.102583.

Kezhen Dong, Hongying Zhang, and Jiangjun Peng. TARGET: Efficient and generalizable multi-view partial multi-label learning via anchor graphs and high-order tensor correlation. *Information Fusion*, 138:104691, 2027. doi: 10.1016/j.inffus.2026.104691.

André Elisseeff and Jason Weston. A kernel method for multi-labelled classification. In *Advances in Neural Information Processing Systems*, volume 14, pp. 681–687, 2001.

Jianqing Fan, Runze Li, Cun-Hui Zhang, and Hui Zou. *Statistical foundations of data science*. CRC press Boca Raton, FL, 2020.

Gene H. Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

Jingqi Liu, Hongying Zhang, Kezhen Dong, and Fei Hao. FastGRAIL: Anchor-graph-based fast adaptive partial multi-label learning with label correlations. *Pattern Recognition*, 178:113457, 2026. doi: 10.1016/j.patcog.2026.113457.

Li-Ping Liu and Thomas G. Dietterich. A conditional multinomial mixture model for superset label learning. In *Advances in Neural Information Processing Systems*, volume 25, pp. 548–556, 2012.

Zewen Liu, Xiaoda Wang, Bohan Wang, Zijie Huang, Carl Yang, and Wei Jin. Graph ODEs and beyond: A comprehensive survey on integrating differential equations with graph neural networks. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6118–6128. ACM, 2025. doi: 10.1145/3711896.3736559.

Colin McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989*, volume 141 of *London Mathematical Society Lecture Note Series*, pp. 148–188. Cambridge University Press, 1989. doi: 10.1017/CBO9781107359949.008.

Charles A. Micchelli and Massimiliano Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, 2005. doi: 10.1162/0899766052530802.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pp. 8024–8035, 2019.

Tiexin Qin, Benjamin Walker, Terry Lyons, Hong Yan, and Haoliang Li. Learning dynamic graph embeddings with neural controlled differential equations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(2):2096–2103, 2026. doi: 10.1109/TPAMI.2025.3617660.

Craig Saunders, Alexander Gammerman, and Volodya Vovk. Ridge regression learning algorithm in dual variables. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 515–521. Morgan Kaufmann Publishers Inc., 1998.

Bernhard Schölkopf and Alexander J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. the MIT Press, 2002.

Konstantinos Trohidis, Grigorios Tsoumakas, George Kalliris, and Ioannis P. Vlahavas. Multi-label classification of music into emotions. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 325–330, 2008.

Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7:91, 2006. doi: 10.1186/1471-2105-7-91.

Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 319–326, 2004.

Jie Wang, Ning Xu, and Xin Geng. VADIS: Investigating inter-view representation biases for multi-view partial multi-label learning. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 4458–4471. PMLR, 2025.

Ziyi Wang and Yitian Xu. A two-stage multi-view partial multi-label learning for enhanced disambiguation. *Knowledge-Based Systems*, 293:111680, 2024. doi: 10.1016/j.knosys.2024.111680.

Jing-Han Wu, Xuan Wu, Qing-Guo Chen, Yao Hu, and Min-Ling Zhang. Feature-induced manifold disambiguation for multi-view partial multi-label learning. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 557–565, 2020.

Yutong Wu, Zheming Xu, Congyan Lang, and Songhe Feng. Integrating cross-graph consensus constraints for label disambiguation in multi-view partial multi-label learning. *Applied Intelligence*, 55:1041, 2025. doi: 10.1007/s10489-025-06889-1.

Louis-Pascal A. C. Xhonneux, Meng Qu, and Jian Tang. Continuous graph neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10432–10441. PMLR, 2020.

Jianhua Xu. Fast multi-label core vector machine. *Pattern Recognition*, 46(3):885–898, 2013.

Ning Xu, Yong-Di Wu, Congyu Qiao, Yi Ren, Minxue Zhang, and Xin Geng. Multi-view partial multi-label learning via graph-fusion-based label enhancement. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11656–11667, 2023. doi: 10.1109/TKDE.2022.3232482.

Fei Yu and Min-Ling Zhang. Maximum margin partial label learning. In *Asian conference on machine learning*, volume 45, pp. 96–111, 2016.

Jiazheng Yuan, Wei Liu, Zhibin Gu, and Songhe Feng. A unified framework for graph-based multi-view partial multi-label learning. *IEEE Access*, 11:49205–49215, 2023. doi: 10.1109/ACCESS.2023.3271730.

Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8):1819–1837, 2014. doi: 10.1109/TKDE.2013.39.

Yongshan Zhang, Jia Wu, Zhihua Cai, and Philip S. Yu. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Transactions on Multimedia*, 22(11):2844–2857, 2020.

Qiyu Zhong, Yi Shan, Haobo Wang, Zhen Yang, and Gengyu Lyu. Tensorized multi-view multi-label classification via Laplace tensor rank. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 78522–78536. PMLR, 2025.

Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2018. doi: 10.1093/nsr/nwx106.

AI USE STATEMENT

Generative AI tools were used in this work for language editing and polishing to improve the readability of the manuscript.

APPENDIX

A ALGORITHMIC DETAILS

Algorithm 1 summarizes the complete training and prediction procedure of GeoCURVE. Given multi-view training data, candidate labels, and capacity parameters, GeoCURVE first constructs view-specific kernels and their shared multi-view geometry. The structural reference is then obtained through capacity-controlled kernel ridge regression, followed by exact leave-one-out estimation to derive structural evidence for candidate reliability estimation. Using the estimated reliability, GeoCURVE constructs reliability-informed supervision and the corresponding candidate-specific correction. The closed-form equilibrium of the reference-restored dynamics is subsequently computed as the training prediction. For unseen samples, GeoCURVE constructs the test–training cross-kernel and obtains the inductive prediction through the corresponding kernel extension. For evaluation, the continuous prediction scores are first normalized by instance-wise min–max scaling and then converted into binary label predictions using a fixed threshold of 0.75.

Algorithm 1 GeoCURVE Training and Prediction

Require: Multi-view training data $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$, candidate-label matrix $\mathbf{Y}$, test samples $\{\mathbf{X}_*^{(v)}\}_{v=1}^{V}$, and capacity parameters γ_r, γ_c

Ensure: Training equilibrium $\mathbf{P}^\star$ and inductive prediction $\mathbf{P}_*^\star$

1: Construct view-specific kernels and the shared kernel $\mathbf{K}$
2: Determine λ_r and λ_c from the prescribed capacities γ_r and γ_c
3: Compute the structural reference $\mathbf{P}_0$ using Eq. (2)
4: Compute exact LOO reference scores using Eq. (3) and obtain pooled rank evidence
5: Estimate the effective negative-like proportion and candidate reliability using Eqs. (5) and (6)
6: Construct the reliability-informed supervision $\widetilde{\mathbf{Y}}$ and candidate-specific correction $\mathbf{R}_{\mathrm{cs}}$ using Eq. (7)
7: Compute the closed-form equilibrium $\mathbf{P}^\star$ using Eq. (12)
8: Construct the shared test–training cross-kernel $\mathbf{K}_*$
9: Compute the inductive prediction $\mathbf{P}_*^\star$ using Eq. (13)

B PROOF OF MODEL PROPERTIES

B.1 VALIDITY OF THE SHARED MULTI-VIEW KERNEL

We show that the fused multi-view kernel used in GeoCURVE corresponds to a valid kernel function on a unified feature space.

Lemma 3. *Let $k^{(v)}$ denote the view-specific Gaussian RBF kernel associated with the v-th view. The fused kernel $\mathbf{K} = \frac{1}{V^2}\sum_{v=1}^{V}\mathbf{K}^{(v)}$ corresponds to a valid kernel function.*

Proof. Since the Gaussian RBF kernel is a valid positive semidefinite kernel, for each view there exists a reproducing kernel Hilbert space $\mathcal{H}_v$ and a feature map $\phi^{(v)}$ such that

$$k^{(v)}(x, x') = \langle \phi^{(v)}(x), \phi^{(v)}(x') \rangle_{\mathcal{H}_v}. \tag{20}$$

We define the direct-sum Hilbert space

$$\mathcal{H} = \bigoplus_{v=1}^{V} \mathcal{H}_v \tag{21}$$

and the fused feature map

$$\phi(x) = \frac{1}{V} \bigoplus_{v=1}^{V} \phi^{(v)}(x). \tag{22}$$

The induced inner product in the fused space satisfies

$$\begin{aligned}\langle \phi(x), \phi(x') \rangle_{\mathcal{H}} &= \frac{1}{V^2} \sum_{v=1}^{V} \langle \phi^{(v)}(x), \phi^{(v)}(x') \rangle_{\mathcal{H}_v} \\ &= \frac{1}{V^2} \sum_{v=1}^{V} k^{(v)}(x, x').\end{aligned} \tag{23}$$

Therefore, the fused kernel is the inner product induced by the direct-sum feature representation. Consequently, it is a valid kernel function, and its Gram matrix is exactly $\mathbf{K} = \frac{1}{V^2} \sum_{v=1}^{V} \mathbf{K}^{(v)}$. □

B.2 Uniqueness of Capacity Parameterization

We prove that the normalized effective capacity provides a well-defined parameterization of the regularization strength used in GeoCURVE.

Theorem 4. *Let $\mathbf{K} \succeq 0$ with $\mathbf{K} \neq \mathbf{0}$ and eigenvalues $\{\mu_s\}_{s=1}^{n}$. Define the normalized effective capacity as*

$$\gamma(\lambda) = \frac{1}{n} \sum_{s=1}^{n} \frac{\mu_s}{\mu_s + \lambda}, \qquad \lambda > 0. \tag{24}$$

For any $\gamma_r, \gamma_c \in (0, \operatorname{rank}(\mathbf{K})/n)$, there exist unique $\lambda_r, \lambda_c > 0$ such that

$$\gamma(\lambda_r) = \gamma_r, \qquad \gamma(\lambda_c) = \gamma_c. \tag{25}$$

Proof. For $\lambda > 0$, the capacity function

$$\gamma(\lambda) = \frac{1}{n} \sum_{s=1}^{n} \frac{\mu_s}{\mu_s + \lambda} \tag{26}$$

is continuous because each term is continuous. For $\mu_s > 0$,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\mu_s}{\mu_s + \lambda} = -\frac{\mu_s}{(\mu_s + \lambda)^2} < 0. \tag{27}$$

For $\mu_s = 0$, the derivative equals zero. Therefore,

$$\gamma'(\lambda) = -\frac{1}{n} \sum_{\mu_s > 0} \frac{\mu_s}{(\mu_s + \lambda)^2} < 0, \tag{28}$$

where the strict inequality follows from the existence of at least one positive eigenvalue of $\mathbf{K}$.

Moreover,

$$\lim_{\lambda \to 0^+} \gamma(\lambda) = \frac{\operatorname{rank}(\mathbf{K})}{n}, \qquad \lim_{\lambda \to \infty} \gamma(\lambda) = 0. \tag{29}$$

Hence, $\gamma(\lambda)$ is continuous and strictly decreasing from $\operatorname{rank}(\mathbf{K})/n$ to 0. Therefore, for every $\gamma \in (0, \operatorname{rank}(\mathbf{K})/n)$, there exists a unique $\lambda > 0$ satisfying the capacity relation. Applying this result to γ_r and γ_c proves the uniqueness of λ_r and λ_c. □

B.3 Self-Influence of the Structural Reference

The structural reference is obtained by the linear smoother $\mathbf{P}_0 = \mathbf{H}_r \mathbf{Y}$. For the i-th instance,

$$\mathbf{P}_{0,i} = \sum_{j=1}^{n} H_{r,ij} \mathbf{Y}_j = H_{r,ii} \mathbf{Y}_i + \sum_{j \neq i} H_{r,ij} \mathbf{Y}_j. \tag{30}$$

The first term represents the direct contribution of the instance's own candidate labels to its reference score. Therefore, using $\mathbf{P}_{0,i}$ to evaluate the reliability of the same instance's candidate labels may introduce self-reinforcement from the observed supervision. The exact leave-one-out prediction removes this direct contribution and is given by Eq. (3).

B.4 Negative-like Mixture Scale Estimation

We provide the theoretical justification for estimating the negative-like mixture contribution from the cumulative evidence distributions.

Lemma 5. *Assume that the candidate evidence distribution satisfies*

$$f_C(z) = \alpha_{\text{eff}} f_N(z) + (1 - \alpha_{\text{eff}}) f_T(z), \tag{31}$$

where $\alpha_{\text{eff}} \in [0, 1]$, $f_N(z)$ denotes the negative-like evidence distribution, and $f_T(z)$ denotes the latent evidence distribution of genuinely relevant candidates. Let G_C, G_N, and G_T denote the corresponding cumulative distributions. Then, for any threshold t satisfying $G_N(t) > 0$,

$$\frac{G_C(t)}{G_N(t)} = \alpha_{\text{eff}} + (1 - \alpha_{\text{eff}}) \frac{G_T(t)}{G_N(t)} \geq \alpha_{\text{eff}}. \tag{32}$$

Moreover, if the contribution of genuinely relevant candidates becomes negligible in low-evidence regions, namely

$$\frac{G_T(t)}{G_N(t)} \to 0, \tag{33}$$

then

$$\frac{G_C(t)}{G_N(t)} \to \alpha_{\text{eff}}. \tag{34}$$

Proof. Integrating the mixture relation over the interval $(-\infty, t]$ gives

$$G_C(t) = \alpha_{\text{eff}} G_N(t) + (1 - \alpha_{\text{eff}}) G_T(t). \tag{35}$$

For any t with $G_N(t) > 0$, dividing both sides by $G_N(t)$ yields

$$\frac{G_C(t)}{G_N(t)} = \alpha_{\text{eff}} + (1 - \alpha_{\text{eff}}) \frac{G_T(t)}{G_N(t)}. \tag{36}$$

Since cumulative distributions are nonnegative,

$$\frac{G_T(t)}{G_N(t)} \geq 0, \tag{37}$$

and therefore

$$\frac{G_C(t)}{G_N(t)} \geq \alpha_{\text{eff}}. \tag{38}$$

If the ratio $G_T(t)/G_N(t)$ approaches zero in low-evidence regions, the second term vanishes and the cumulative ratio converges to α_{eff}. □

B.5 Global Well-Posedness and Exponential Convergence

Theorem 6. *The dynamics in Eq. (10) admit a unique global solution and converge exponentially to the unique equilibrium $\mathbf{P}^\star$ in Eq. (12), with*

$$\|\mathbf{P}(t) - \mathbf{P}^\star\|_F \leq e^{-\xi_c t} \|\mathbf{P}(0) - \mathbf{P}^\star\|_F. \tag{39}$$

Proof. Let $\mathbf{R}_{\text{cs}} = \widetilde{\mathbf{Y}} - \mathbf{P}_0$. The reference-restored dynamics in Eq. (10) can be rewritten as

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = -\mathbf{A}_c\big(\mathbf{P}(t) - \mathbf{P}_0\big) + \overline{\mathbf{K}}\mathbf{R}_{\text{cs}}, \tag{40}$$

where

$$\mathbf{A}_c = \overline{\mathbf{K}} + \xi_c \mathbf{I}. \tag{41}$$

Since $\mathbf{K} \succeq 0$, we have $\overline{\mathbf{K}} \succeq 0$. Together with $\xi_c > 0$, $\mathbf{A}_c$ is symmetric positive definite.

The vector field of the above affine system is globally Lipschitz. For any two prediction states $\mathbf{P}_1$ and $\mathbf{P}_2$,

$$\|F(\mathbf{P}_1) - F(\mathbf{P}_2)\|_F = \|\mathbf{A}_c(\mathbf{P}_1 - \mathbf{P}_2)\|_F \leq \|\mathbf{A}_c\|_2 \|\mathbf{P}_1 - \mathbf{P}_2\|_F. \tag{42}$$

Therefore, the initial-value problem admits a unique global solution.

An equilibrium $\mathbf{P}^\star$ satisfies

$$\mathbf{A}_c(\mathbf{P}^\star - \mathbf{P}_0) = \overline{\mathbf{K}}\mathbf{R}_{\mathrm{cs}}. \tag{43}$$

Since $\mathbf{A}_c$ is nonsingular, the equilibrium is unique and given by

$$\mathbf{P}^\star = \mathbf{P}_0 + \mathbf{A}_c^{-1}\overline{\mathbf{K}}\mathbf{R}_{\mathrm{cs}}. \tag{44}$$

Using $\overline{\mathbf{K}} = \mathbf{K}/\mu_{\max}$ and $\lambda_c = \xi_c\mu_{\max}$, we have

$$\mathbf{A}_c^{-1}\overline{\mathbf{K}} = (\overline{\mathbf{K}} + \xi_c\mathbf{I})^{-1}\overline{\mathbf{K}} = (\mathbf{K} + \lambda_c\mathbf{I})^{-1}\mathbf{K} = \mathbf{K}(\mathbf{K} + \lambda_c\mathbf{I})^{-1}. \tag{45}$$

Therefore,

$$\mathbf{P}^\star = \mathbf{P}_0 + \mathbf{K}(\mathbf{K} + \lambda_c\mathbf{I})^{-1}\mathbf{R}_{\mathrm{cs}}, \tag{46}$$

which is exactly Eq. (12).

Subtracting the equilibrium equation from the dynamics gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathbf{P}(t) - \mathbf{P}^\star\big) = -\mathbf{A}_c\big(\mathbf{P}(t) - \mathbf{P}^\star\big). \tag{47}$$

The solution is

$$\mathbf{P}(t) - \mathbf{P}^\star = e^{-t\mathbf{A}_c}\big(\mathbf{P}(0) - \mathbf{P}^\star\big). \tag{48}$$

Let ν_s denote the eigenvalues of $\overline{\mathbf{K}}$. Since $\mathbf{A}_c = \overline{\mathbf{K}} + \xi_c\mathbf{I}$, the eigenvalues of $\mathbf{A}_c$ are $\nu_s + \xi_c$. Therefore,

$$\lambda_{\min}(\mathbf{A}_c) \geq \xi_c. \tag{49}$$

Because $\mathbf{A}_c$ is symmetric positive definite,

$$\|e^{-t\mathbf{A}_c}\|_2 = e^{-t\lambda_{\min}(\mathbf{A}_c)} \leq e^{-\xi_c t}. \tag{50}$$

Combining the above results gives

$$\|\mathbf{P}(t) - \mathbf{P}^\star\|_F \leq e^{-\xi_c t}\|\mathbf{P}(0) - \mathbf{P}^\star\|_F, \tag{51}$$

which proves the global exponential convergence. □

B.6 Spectral Characterization of the GeoCURVE Equilibrium

Proposition 7. *Let $\mathbf{K} = \mathbf{U}\,\mathrm{diag}(\mu_1, \ldots, \mu_n)\mathbf{U}^T$. For any matrix $\mathbf{A} \in \mathbb{R}^{n\times C}$, denote its component along the s-th eigendirection by $\mathbf{A}_s^{\mathrm{sp}} = \mathbf{u}_s^T\mathbf{A}$. Then the equilibrium satisfies*

$$\mathbf{P}_s^{\star,\mathrm{sp}} = (1 - g_s)\mathbf{P}_{0,s}^{\mathrm{sp}} + g_s\widetilde{\mathbf{Y}}_s^{\mathrm{sp}}, \tag{52}$$

where $g_s = \mu_s/(\mu_s + \lambda_c)$ is the correction retention coefficient. Moreover, $0 \leq g_s < 1$, g_s strictly increases with μ_s, and for $\mu_s > 0$, g_s strictly decreases with λ_c.

Proof. From Eq. (12),

$$\mathbf{P}^\star = \mathbf{P}_0 + \mathbf{H}_c\mathbf{R}_{\mathrm{cs}}, \qquad \mathbf{H}_c = \mathbf{K}(\mathbf{K} + \lambda_c\mathbf{I})^{-1}. \tag{53}$$

Using $\mathbf{K} = \mathbf{U}\,\mathrm{diag}(\mu_1, \ldots, \mu_n)\mathbf{U}^T$, we obtain

$$\mathbf{H}_c = \mathbf{U}\,\mathrm{diag}\Big(\frac{\mu_1}{\mu_1 + \lambda_c}, \ldots, \frac{\mu_n}{\mu_n + \lambda_c}\Big)\mathbf{U}^T. \tag{54}$$

Left-multiplying the equilibrium by $\mathbf{u}_s^T$ gives

$$\mathbf{P}_s^{\star,\mathrm{sp}} = \mathbf{P}_{0,s}^{\mathrm{sp}} + \frac{\mu_s}{\mu_s + \lambda_c}\mathbf{R}_{\mathrm{cs},s}^{\mathrm{sp}}. \tag{55}$$

Using $\mathbf{R}_{\text{cs}} = \widetilde{\mathbf{Y}} - \mathbf{P}_0$, we further obtain

$$\mathbf{P}_s^{\star,\text{sp}} = \frac{\lambda_c}{\mu_s + \lambda_c}\mathbf{P}_{0,s}^{\text{sp}} + \frac{\mu_s}{\mu_s + \lambda_c}\widetilde{\mathbf{Y}}_s^{\text{sp}}. \tag{56}$$

Defining $g_s = \mu_s/(\mu_s + \lambda_c)$ yields

$$\mathbf{P}_s^{\star,\text{sp}} = (1 - g_s)\mathbf{P}_{0,s}^{\text{sp}} + g_s\widetilde{\mathbf{Y}}_s^{\text{sp}}. \tag{57}$$

Since $\mu_s \geq 0$ and $\lambda_c > 0$, $0 \leq g_s < 1$. Moreover,

$$\frac{\partial g_s}{\partial \mu_s} = \frac{\lambda_c}{(\mu_s + \lambda_c)^2} > 0, \qquad \frac{\partial g_s}{\partial \lambda_c} = -\frac{\mu_s}{(\mu_s + \lambda_c)^2} \leq 0. \tag{58}$$

Therefore, g_s strictly increases with μ_s and, for $\mu_s > 0$, strictly decreases with λ_c. □

C Proofs of Theoretical Results

C.1 Proof of Theorem 1

Proof. Recall that the GeoCURVE equilibrium can be written as

$$\mathbf{P}^\star = (\mathbf{I} - \mathbf{H}_c)\mathbf{P}_0 + \mathbf{H}_c\widetilde{\mathbf{Y}}, \qquad \mathbf{P}_0 = \mathbf{H}_r\mathbf{Y}. \tag{59}$$

For the oracle construction, $\mathbf{P}_0^{\text{or}} = \mathbf{H}_r\mathbf{Y}^{\text{true}}$, and hence

$$\mathbf{P}_{\text{or}}^\star = (\mathbf{I} - \mathbf{H}_c)\mathbf{P}_0^{\text{or}} + \mathbf{H}_c\mathbf{Y}^{\text{true}}. \tag{60}$$

Subtracting Eq. (60) from Eq. (59) gives

$$\mathbf{P}^\star - \mathbf{P}_{\text{or}}^\star = (\mathbf{I} - \mathbf{H}_c)\mathbf{H}_r(\mathbf{Y} - \mathbf{Y}^{\text{true}}) + \mathbf{H}_c(\widetilde{\mathbf{Y}} - \mathbf{Y}^{\text{true}}). \tag{61}$$

Taking Frobenius norms and using $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2\|\mathbf{B}\|_F$ yields

$$\|\mathbf{P}^\star - \mathbf{P}_{\text{or}}^\star\|_F \leq \|(\mathbf{I} - \mathbf{H}_c)\mathbf{H}_r\|_2\|\mathbf{Y} - \mathbf{Y}^{\text{true}}\|_F + \|\mathbf{H}_c\|_2\|\widetilde{\mathbf{Y}} - \mathbf{Y}^{\text{true}}\|_F. \tag{62}$$

Let $\mathbf{K} = \mathbf{U}\,\text{diag}(\mu_1, \ldots, \mu_n)\mathbf{U}^T$. Since

$$\mathbf{H}_r = \mathbf{U}\,\text{diag}\big(\frac{\mu_s}{\mu_s + \lambda_r}\big)\mathbf{U}^T \tag{63}$$

and

$$\mathbf{H}_c = \mathbf{U}\,\text{diag}\big(\frac{\mu_s}{\mu_s + \lambda_c}\big)\mathbf{U}^T, \tag{64}$$

the two operators share the same eigenvectors. Therefore,

$$(\mathbf{I} - \mathbf{H}_c)\mathbf{H}_r = \mathbf{U}\,\text{diag}\big(\frac{\lambda_c\mu_s}{(\mu_s + \lambda_c)(\mu_s + \lambda_r)}\big)\mathbf{U}^T. \tag{65}$$

It follows that

$$\|(\mathbf{I} - \mathbf{H}_c)\mathbf{H}_r\|_2 = \max_s \frac{\lambda_c\mu_s}{(\mu_s + \lambda_c)(\mu_s + \lambda_r)} = \rho_{r,c}. \tag{66}$$

For every $\mu_s \geq 0$,

$$(\mu_s + \lambda_c)(\mu_s + \lambda_r) - \lambda_c\mu_s = \mu_s^2 + \lambda_r\mu_s + \lambda_c\lambda_r > 0, \tag{67}$$

because $\lambda_r > 0$ and $\lambda_c > 0$. Hence,

$$0 \leq \rho_{r,c} < 1. \tag{68}$$

Similarly,

$$\|\mathbf{H}_c\|_2 = \max_s \frac{\mu_s}{\mu_s + \lambda_c} = \frac{\mu_{\max}}{\mu_{\max} + \lambda_c} = \vartheta_c. \tag{69}$$

Using $\lambda_c = \xi_c \mu_{\max}$ gives

$$\vartheta_c = \frac{1}{1+\xi_c} < 1. \tag{70}$$

Substituting these two operator-norm bounds into Eq. (62) yields

$$\|\mathbf{P}^\star - \mathbf{P}^\star_{\mathrm{or}}\|_F \le \rho_{r,c}\|\mathbf{Y} - \mathbf{Y}^{\mathrm{true}}\|_F + \vartheta_c\|\widetilde{\mathbf{Y}} - \mathbf{Y}^{\mathrm{true}}\|_F, \tag{71}$$

which proves Eq. (14).

Finally, by the triangle inequality,

$$\begin{aligned}\|\mathbf{P}^\star - \mathbf{Y}^{\mathrm{true}}\|_F &\le \|\mathbf{P}^\star - \mathbf{P}^\star_{\mathrm{or}}\|_F + \|\mathbf{P}^\star_{\mathrm{or}} - \mathbf{Y}^{\mathrm{true}}\|_F \\ &\le \|\mathbf{P}^\star_{\mathrm{or}} - \mathbf{Y}^{\mathrm{true}}\|_F + \rho_{r,c}\|\mathbf{Y} - \mathbf{Y}^{\mathrm{true}}\|_F + \vartheta_c\|\widetilde{\mathbf{Y}} - \mathbf{Y}^{\mathrm{true}}\|_F,\end{aligned} \tag{72}$$

which proves Eq. (15) and completes the proof. □

C.2 Proof of Theorem 2

Lemma 8. *Let $\mathbf{h}_c$ denote the candidate-specific correction function induced by GeoCURVE, defined as*

$$\mathbf{h}_c(x) = \mathbf{k}_x^T(\mathbf{K} + \lambda_c\mathbf{I})^{-1}\mathbf{R}_{\mathrm{cs}}, \tag{73}$$

where $\mathbf{k}_x = [k(x_1,x),\dots,k(x_n,x)]^T$. Let $\mathbf{K} = \mathbf{U}\,\mathrm{diag}(\mu_1,\dots,\mu_n)\mathbf{U}^T$, and define the correction component along the s-th kernel eigendirection as

$$\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s} = \mathbf{u}_s^T\mathbf{R}_{\mathrm{cs}}, \tag{74}$$

where $\mathbf{u}_s$ is the s-th eigenvector of $\mathbf{K}$. Then its squared RKHS norm satisfies

$$\|\mathbf{h}_c\|^2_{\mathcal{H}_k^C} = \sum_{\mu_s>0} \frac{\mu_s}{(\mu_s+\lambda_c)^2}\|\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s}\|_2^2. \tag{75}$$

Proof. Let

$$\mathbf{B}_c = (\mathbf{K} + \lambda_c\mathbf{I})^{-1}\mathbf{R}_{\mathrm{cs}}. \tag{76}$$

Then $\mathbf{h}_c(x) = \mathbf{k}_x^T\mathbf{B}_c$.

For the direct-sum vector-valued RKHS $\mathcal{H}_k^C = \mathcal{H}_k \oplus \cdots \oplus \mathcal{H}_k$, let $\mathbf{b}_j$ denote the j-th column of $\mathbf{B}_c$. Since $h_{c,j}(x) = \sum_{i=1}^n B_{c,ij}k(x_i,x)$, the reproducing property gives $\|h_{c,j}\|^2_{\mathcal{H}_k} = \mathbf{b}_j^T\mathbf{K}\mathbf{b}_j$ (Micchelli & Pontil, 2005). Therefore,

$$\|\mathbf{h}_c\|^2_{\mathcal{H}_k^C} = \sum_{j=1}^C \mathbf{b}_j^T\mathbf{K}\mathbf{b}_j = \mathrm{tr}(\mathbf{B}_c^T\mathbf{K}\mathbf{B}_c). \tag{77}$$

Substituting $\mathbf{B}_c$ gives

$$\|\mathbf{h}_c\|^2_{\mathcal{H}_k^C} = \mathrm{tr}(\mathbf{R}_{\mathrm{cs}}^T(\mathbf{K}+\lambda_c\mathbf{I})^{-1}\mathbf{K}(\mathbf{K}+\lambda_c\mathbf{I})^{-1}\mathbf{R}_{\mathrm{cs}}). \tag{78}$$

Using the eigendecomposition of $\mathbf{K}$, we have

$$(\mathbf{K}+\lambda_c\mathbf{I})^{-1}\mathbf{K}(\mathbf{K}+\lambda_c\mathbf{I})^{-1} = \mathbf{U}\,\mathrm{diag}\Big(\frac{\mu_1}{(\mu_1+\lambda_c)^2},\dots,\frac{\mu_n}{(\mu_n+\lambda_c)^2}\Big)\mathbf{U}^T. \tag{79}$$

Therefore, using $\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s} = \mathbf{u}_s^T\mathbf{R}_{\mathrm{cs}}$, we obtain

$$\|\mathbf{h}_c\|^2_{\mathcal{H}_k^C} = \sum_{\mu_s>0} \frac{\mu_s}{(\mu_s+\lambda_c)^2}\|\mathbf{R}^{\mathrm{sp}}_{\mathrm{cs},s}\|_2^2. \tag{80}$$

□

Lemma 9. *Let $Z_1, \dots, Z_n$ be i.i.d. random elements in a Hilbert space $\mathcal{H}$ satisfying $\|Z_i\|_{\mathcal{H}} \le B$ almost surely. Then, for any $\eta \in (0,1)$, with probability at least $1-\eta$,*

$$\|\frac{1}{n}\sum_{i=1}^{n} Z_i - \mathbb{E}Z_1\|_{\mathcal{H}} \le \frac{B}{\sqrt{n}}(1+\sqrt{2\log(1/\eta)}). \tag{81}$$

Proof. Let $\overline{Z} = n^{-1}\sum_{i=1}^{n} Z_i$ and $\mu = \mathbb{E}Z_1$. By Jensen's inequality and independence,

$$\mathbb{E}\|\overline{Z}-\mu\|_{\mathcal{H}} \le \sqrt{\mathbb{E}\|\overline{Z}-\mu\|_{\mathcal{H}}^2} = \sqrt{\frac{1}{n}\mathbb{E}\|Z_1-\mu\|_{\mathcal{H}}^2} \le \frac{B}{\sqrt{n}}. \tag{82}$$

Now define $F(Z_1,\dots,Z_n) = \|\overline{Z}-\mu\|_{\mathcal{H}}$. Replacing one Z_i by Z_i' changes F by at most $\|Z_i - Z_i'\|_{\mathcal{H}}/n \le 2B/n$. Hence, by McDiarmid's bounded-difference inequality (McDiarmid, 1989), with probability at least $1-\eta$,

$$F \le \mathbb{E}F + B\sqrt{\frac{2\log(1/\eta)}{n}}. \tag{83}$$

Combining this inequality with Eq. (82) proves Eq. (81). □

Proof of Theorem 2. By Lemma 8, Eq. (17) holds. It remains to establish the generalization bound.

Let $\phi(x) \in \mathcal{H}_k$ denote a feature representation of the fixed kernel k, so that $k(x,x') = \langle \phi(x), \phi(x')\rangle_{\mathcal{H}_k}$. Since $k(x,x) \le \kappa^2$, $\|\phi(x)\|_{\mathcal{H}_k} \le \kappa$.

For a vector-valued function $\mathbf{f} = (f_1,\dots,f_C) \in \mathcal{H}_k^C$, write $f_c(x) = \langle w_c, \phi(x)\rangle_{\mathcal{H}_k}$, and let $\mathbf{W} = (w_1,\dots,w_C)$. Then $\|\mathbf{f}\|_{\mathcal{H}_k^C} = \|\mathbf{W}\|_{\mathcal{H}_k^C}$.

Define the population and empirical second-moment operators

$$\mathbf{\Sigma} = \mathbb{E}[\phi(X)\otimes\phi(X)], \qquad \widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\phi(X_i)\otimes\phi(X_i), \tag{84}$$

and the corresponding label-feature moments

$$\mathbf{m} = \big(\mathbb{E}[Y_c^{\mathrm{true}}\phi(X)]\big)_{c=1}^{C}, \qquad \widehat{\mathbf{m}} = \frac{1}{n}\sum_{i=1}^{n}\big(Y_{ic}^{\mathrm{true}}\phi(X_i)\big)_{c=1}^{C}. \tag{85}$$

We let $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ act componentwise on $\mathcal{H}_k^C$.

Let $\mathbf{W}_0$ and $\mathbf{W}_h$ represent $\mathbf{f}_0$ and $\mathbf{h}$, respectively, with $q_0 = \|\mathbf{W}_0\|_{\mathcal{H}_k^C}$ and $q = \|\mathbf{W}_h\|_{\mathcal{H}_k^C}$. Using the squared-loss risks defined in Eq. (16), the population risk increment can be written as

$$\Delta\mathcal{R}(\mathbf{h}) = \frac{1}{C}\Big[2\langle \mathbf{W}_0, \mathbf{\Sigma}\mathbf{W}_h\rangle + \langle \mathbf{W}_h, \mathbf{\Sigma}\mathbf{W}_h\rangle - 2\langle \mathbf{W}_h, \mathbf{m}\rangle\Big], \tag{86}$$

and $\Delta\widehat{\mathcal{R}}(\mathbf{h})$ has the same form with $\mathbf{\Sigma}$ and $\mathbf{m}$ replaced by $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{m}}$. Using Cauchy–Schwarz and the operator-norm inequality $|\langle u, Av\rangle| \le \|u\|\,\|A\|_{\mathrm{op}}\,\|v\|$, we have

$$\big|\Delta\mathcal{R}(\mathbf{h}) - \Delta\widehat{\mathcal{R}}(\mathbf{h})\big| \le \frac{1}{C}\Big[(2q_0q+q^2)\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\|_{\mathrm{op}} + 2q\|\mathbf{m}-\widehat{\mathbf{m}}\|_{\mathcal{H}_k^C}\Big]. \tag{87}$$

We next bound the two sample moments. Since $\|\phi(X)\otimes\phi(X)\|_{\mathrm{HS}} = \|\phi(X)\|_{\mathcal{H}_k}^2 \le \kappa^2$, applying Lemma 9 in the Hilbert–Schmidt operator space with $B = \kappa^2$ and $\eta = \delta/2$ gives, with probability at least $1-\delta/2$,

$$\|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\|_{\mathrm{op}} \le \|\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}\|_{\mathrm{HS}} \le \kappa^2 a_n(\delta), \tag{88}$$

where

$$a_n(\delta) = \frac{1+\sqrt{2\log(2/\delta)}}{\sqrt{n}}. \tag{89}$$

Moreover, $\|\mathbf{Y}^{\mathrm{true}}\|_2 \le \sqrt{C}$ and $\|\phi(X)\|_{\mathcal{H}_k} \le \kappa$, so

$$\|(Y_c^{\mathrm{true}}\phi(X))_{c=1}^{C}\|_{\mathcal{H}_k^C} \le \kappa\sqrt{C}. \tag{90}$$

Applying Lemma 9 in $\mathcal{H}_k^C$ with $B = \kappa\sqrt{C}$ and $\eta = \delta/2$ gives

$$\|\mathbf{m} - \widehat{\mathbf{m}}\|_{\mathcal{H}_k^C} \le \kappa\sqrt{C}\, a_n(\delta). \tag{91}$$

By a union bound, both inequalities hold simultaneously with probability at least $1-\delta$. Substituting them into Eq. (87) yields

$$\left|\Delta\mathcal{R}(\mathbf{h}) - \Delta\widehat{\mathcal{R}}(\mathbf{h})\right| \le \frac{a_n(\delta)}{C}\Big[\kappa^2 q(2q_0+q) + 2\kappa\sqrt{C}\, q\Big]. \tag{92}$$

The concentration event above depends only on $\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}$ and $\widehat{\mathbf{m}} - \mathbf{m}$, and does not depend on the particular choice of $\mathbf{f}_0$ or $\mathbf{h}$. Hence, on this event, Eq. (87) holds simultaneously for all finite-norm $\mathbf{f}_0, \mathbf{h} \in \mathcal{H}_k^C$, and therefore also for the data-dependent functions produced by GeoCURVE. Hence, on this event, Eq. (87) holds simultaneously for all finite-norm $\mathbf{f}_0, \mathbf{h} \in \mathcal{H}_k^C$. It therefore remains valid for the data-dependent functions produced by GeoCURVE. Taking $\mathbf{h} = \mathbf{h}_c$ and $q = q_c$ gives

$$\left|\Delta\mathcal{R}(\mathbf{h}_c) - \Delta\widehat{\mathcal{R}}(\mathbf{h}_c)\right| \le \Psi_n(q_0, q_c, \delta), \tag{93}$$

where

$$\Psi_n(q_0, q_c, \delta) = \frac{a_n(\delta)}{C}\Big[\kappa^2 q_c(2q_0+q_c) + 2\kappa\sqrt{C}\, q_c\Big]. \tag{94}$$

Finally, for $q_c \ge 0$,

$$\frac{\partial\Psi_n(q_0, q_c, \delta)}{\partial q_c} = \frac{2a_n(\delta)}{C}\Big[\kappa^2(q_0+q_c) + \kappa\sqrt{C}\Big] \ge 0. \tag{95}$$

Hence, $\Psi_n(q_0, q_c, \delta)$ is nondecreasing in q_c. This completes the proof. □

D Experimental Details

D.1 Dataset Details

Table 3 summarizes the statistics of the six benchmark datasets used in our experiments. The reported dimensions follow the actual view order used by the experimental implementation.

Table 3: Statistics of the experimental datasets, where n, V, and C denote the numbers of instances, views, and labels, respectively.

Dataset	n	V	C	View dimensions
Emotions	593	2	6	8/64
Yeast	2417	2	14	79/24
EspGame5k	4957	4	60	48/91/519/368
OBJECT	6047	5	31	64/225/144/73/128
Human	3106	3	14	20/20/400
Plant	978	3	14	20/20/400

D.2 Compared Methods and Parameter Settings

We compare GeoCURVE with six representative competing methods. Unless otherwise stated, the parameter configurations follow the recommendations in the corresponding papers or publicly available implementations.

- **FastGRAIL** (Liu et al., 2026): We adopt the original parameter setting with $\gamma = 1$, $\lambda = 1$, $\eta = 0.1$, $\rho = 100$, $m = \lfloor\sqrt{n}/2\rfloor$, and $k = 10$.
- **TMvML** (Zhong et al., 2025): The parameters are configured as $\alpha \in \{10^{-1}, 10^0, \ldots, 10^5\}$, $\beta \in \{10^{-5}, 10^{-4}, \ldots, 10^0\}$, and $\delta \in \{10^{-5}, 10^{-4}, \ldots, 10^0\}$.

- **GRADIS** (Chen et al., 2020): We use the recommended setting $k = 8$, $\alpha = 0.95$, $\gamma = 0.1$, and $\eta = 0.1$.
- **FIMAN** (Wu et al., 2020): The parameters are set to $t_d = 0.4$, $t_p = 0.6$, $k = 10$, and $\eta = 1$.
- **GLADE** (Xu et al., 2023): The parameters are configured as $\lambda = 0.001$, $\gamma_1 = 65$, and $\gamma_2 = 50$.
- **MSPML-TLR** (Deng et al., 2024): We set $\alpha = 0.2$, $\beta = 0.4$, $\gamma = 0.1$, $\delta = 40$, $\lambda_1 = 0.5$, $\lambda_2 = 1$, and $c = 5$.
- **VADIS** (Wang et al., 2025): The parameters are set as $\gamma_1 = 5$, $\gamma_2 = 17.79$, and $\gamma_3 = 17.79$. The maximum number of optimization iterations is set to 30, with 300 BFGS iterations and 100 ADMM iterations.

D.3 Implementation and Hyperparameter-Selection Details

Feature Preprocessing and Kernel Construction. All preprocessing statistics and kernel parameters are estimated exclusively from the training portion of each split and then applied unchanged to the corresponding validation or test data. For each view, features are standardized using the training-set mean and standard deviation, followed by instance-wise ℓ_2 normalization. The RBF scale σ_v^2 is set to the median positive off-diagonal squared Euclidean distance among the normalized training samples. The resulting view-specific kernels are fused as $\mathbf{K} = V^{-2} \sum_{v=1}^{V} \mathbf{K}^{(v)}$, and the same training-derived RBF scales are used to construct the validation–training and test–training cross-kernels.

Nested Hyperparameter Selection. For each outer split, one fold is held out for final testing, while the remaining nine folds form the outer-training set used for model fitting and hyperparameter selection. The nine outer-training folds are rotated to form nine inner splits, with eight folds used for fitting and one for validation. Feature preprocessing, RBF-scale estimation, and kernel construction are recomputed using only the eight inner-training folds. Hence, hyperparameter selection remains entirely within the outer-training set and never accesses the underlying clean labels.

For each inner split, a pilot reference model with $\gamma_r = 0.1$ is fitted using only the eight inner-training folds and then applied to the held-out validation fold. A fixed weak validation target is subsequently constructed from the resulting reference predictions and the observed candidate annotations. The same target is used for all candidate pairs (γ_r, γ_c) within that split. Let $\widetilde{\mathbf{Y}}^{\mathrm{val}}$ denote this reliability-informed target and $\mathbf{S}^{\mathrm{val}}$ the corresponding prediction scores. We evaluate each candidate pair using a soft extension of average precision, obtained by replacing binary relevance indicators with $\widetilde{\mathbf{Y}}^{\mathrm{val}}$:

$$\mathrm{SoftAP}_i = \frac{\displaystyle\sum_{k=1}^{C} \widetilde{Y}^{\mathrm{val}}_{i,\pi_i(k)} \, k^{-1} \sum_{\ell=1}^{k} \widetilde{Y}^{\mathrm{val}}_{i,\pi_i(\ell)}}{\displaystyle\sum_{j=1}^{C} \widetilde{Y}^{\mathrm{val}}_{ij}}, \tag{96}$$

where $\pi_i(k)$ denotes the label ranked at position k for validation instance i. Instances with zero target mass are excluded.

The inner-fold SoftAP values are aggregated according to the number of valid validation instances, and the pair (γ_r, γ_c) with the largest aggregated value is selected. Exact ties are resolved by choosing the smaller γ_r and then the smaller γ_c. GeoCURVE is finally refitted on all nine outer-training folds using the selected capacities and evaluated once on the untouched outer test fold.

Candidate-Reliability Estimation. For a fixed γ_r, GeoCURVE computes the structural reference and its exact KRR leave-one-out scores under the fixed smoother. The pooled scores are transformed into normalized mid-rank coordinates, from which candidate and trusted-negative evidence densities are estimated separately using Gaussian-smoothed histograms with 128 bins and a robust Silverman bandwidth. The effective negative-like mixture scale is estimated from the minimum left-tail empirical CDF ratio over $\mathcal{T} = \{0.05, 0.075, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50\}$, and is combined with

the density ratio to compute the candidate reliability in Eq. (6). All reliability-estimation settings are fixed throughout the experiments and are never selected using clean labels.

Capacity Solution and Prediction. For each selected normalized capacity $\gamma \in \{\gamma_r, \gamma_c\}$, the corresponding regularization parameter λ is obtained by solving $\gamma = n^{-1}\sum_i \mu_i/(\mu_i + \lambda)$ through one-dimensional bisection. This gives λ_r for the structural reference and λ_c for the correction operator.

Training predictions are obtained from the closed-form equilibrium induced by the continuous-time dynamics, avoiding numerical integration while preserving the dynamical interpretation. For unseen samples, prediction uses the corresponding inductive extension based on the test–training cross-kernel and quantities estimated exclusively from the outer-training fold.

Evaluation and Numerical Settings. All GeoCURVE experiments use double-precision floating-point arithmetic and a fixed random seed of 42. Average precision, one-error, ranking loss, and coverage are computed directly from the continuous prediction scores.

For Hamming loss, the prediction scores of each test instance are first independently min–max normalized to obtain $\widetilde{\mathbf{P}}^{\star}$, and then binarized as

$$\widehat{Y}_{ij} = \mathbf{1}\big(\widetilde{P}^{\star}_{ij} > 0.75\big). \tag{97}$$

The threshold 0.75 is fixed across all datasets and ambiguity configurations.

Computing Environment. All experiments were implemented in PyTorch (Paszke et al., 2019) and executed on a server equipped with four NVIDIA RTX 4090 GPUs, with each run using a single GPU.

D.4 Complete Comparison Results

The main text reports the complete average precision results under all dataset–ambiguity configurations. For completeness, Tables 4–7 report the corresponding results in terms of ranking loss, coverage, Hamming loss, and one-error. For all four metrics, lower values indicate better performance. The best and second-best results in each configuration are shown in bold and underlined, respectively.

GeoCURVE exhibits consistently strong performance across these complementary metrics. It ranks first overall on RL, COV, and OE, placing among the top two methods in 36, 36, and 35 of the 36 cases, respectively. On HL, GeoCURVE ranks second overall and achieves top-two performance in 31 of the 36 cases. The comparatively weaker HL results may partly reflect the difference between the model-selection criterion and the evaluation metric: the capacities are selected using SoftAP in inner validation, whereas HL relies on thresholded binary predictions. Overall, these results further demonstrate the effectiveness and consistent performance of GeoCURVE across different evaluation criteria.

D.5 Statistical Significance Analysis

We assess differences among the compared methods using nonparametric statistical tests (Demšar, 2006). Each dataset under one ambiguity configuration is treated as a benchmark case, giving $N = 36$ cases and $K = 8$ methods. Methods are ranked independently within each case, with rank 1 assigned to the best method and average ranks used for ties. Larger values are preferred for AP, whereas smaller values are preferred for HL, OE, RL, and COV. All tests use the original unrounded mean results.

We first apply the Friedman test with the Iman–Davenport correction. Table 8 shows that the F_F statistics substantially exceed the critical value at $\alpha = 0.05$ for all five metrics, rejecting the null hypothesis of equal performance. We then conduct the Bonferroni–Dunn post-hoc test with GeoCURVE as the control method. The critical-difference diagrams in Fig. 3 show that GeoCURVE achieves the best average rank on AP, OE, RL, and COV and the second-best rank on HL. GeoCURVE significantly outperforms all competitors on RL and COV; its differences from GRADIS on AP and OE, and from GRADIS and FastGRAIL on HL, are not statistically significant.

Table 4: Ranking loss (RL ↓) results under different ambiguity configurations. The best and second-best results in each configuration are shown in bold and underlined, respectively.

Method	(p, r)	Emotions	EspGame5k	Yeast	OBJECT	Human	Plant
FastGRAIL	$(0.3, 1)$	0.172±0.029	0.210±0.007	0.174±0.009	0.090±0.006	0.169±0.008	0.170±0.017
	$(0.3, 2)$	0.181±0.026	0.213±0.007	0.175±0.009	0.092±0.006	0.172±0.008	0.183±0.009
	$(0.3, 3)$	0.178±0.023	0.217±0.006	0.177±0.010	0.093±0.007	0.178±0.011	0.185±0.012
	$(0.7, 1)$	0.184±0.017	0.215±0.007	0.176±0.009	0.092±0.006	0.175±0.009	0.185±0.017
	$(0.7, 2)$	0.184±0.025	0.223±0.009	0.178±0.009	0.095±0.006	0.180±0.010	0.191±0.020
	$(0.7, 3)$	0.197±0.024	0.227±0.008	0.181±0.010	0.100±0.006	0.189±0.013	0.203±0.012
TMvML	$(0.3, 1)$	0.186±0.024	0.189±0.007	0.179±0.011	0.124±0.004	0.151±0.010	0.165±0.015
	$(0.3, 2)$	0.199±0.027	0.192±0.008	0.183±0.011	0.124±0.006	0.163±0.005	0.178±0.018
	$(0.3, 3)$	0.201±0.030	0.194±0.007	0.186±0.012	0.128±0.005	0.162±0.011	0.176±0.023
	$(0.7, 1)$	0.188±0.018	0.192±0.005	0.185±0.010	0.126±0.006	0.161±0.007	0.178±0.020
	$(0.7, 2)$	0.208±0.029	0.205±0.006	0.187±0.012	0.128±0.009	0.168±0.011	0.194±0.017
	$(0.7, 3)$	0.211±0.023	0.202±0.007	0.191±0.014	0.133±0.007	0.181±0.010	0.200±0.015
GRADIS	$(0.3, 1)$	0.150±0.028	0.206±0.005	0.168±0.010	0.099±0.004	0.155±0.013	0.160±0.011
	$(0.3, 2)$	**0.145±0.021**	0.211±0.008	0.169±0.010	0.102±0.005	0.155±0.013	0.174±0.016
	$(0.3, 3)$	**0.149±0.020**	0.215±0.006	0.172±0.011	0.104±0.004	0.156±0.013	0.177±0.017
	$(0.7, 1)$	**0.144±0.030**	0.212±0.008	0.171±0.012	0.105±0.006	0.156±0.012	0.176±0.018
	$(0.7, 2)$	0.175±0.025	0.224±0.005	0.179±0.013	0.112±0.006	0.159±0.009	0.188±0.016
	$(0.7, 3)$	0.195±0.032	0.234±0.006	0.184±0.013	0.115±0.007	0.166±0.013	0.182±0.011
FIMAN	$(0.3, 1)$	0.175±0.022	0.225±0.007	0.183±0.011	0.097±0.006	0.187±0.011	0.257±0.029
	$(0.3, 2)$	0.179±0.023	0.229±0.007	0.185±0.012	0.099±0.006	0.194±0.012	0.272±0.036
	$(0.3, 3)$	0.176±0.024	0.233±0.006	0.185±0.011	0.099±0.007	0.194±0.014	0.272±0.019
	$(0.7, 1)$	0.186±0.025	0.229±0.007	0.186±0.012	0.101±0.005	0.202±0.013	0.285±0.026
	$(0.7, 2)$	0.194±0.016	0.241±0.009	0.186±0.013	0.104±0.006	0.220±0.013	0.296±0.022
	$(0.7, 3)$	**0.188±0.026**	0.243±0.009	0.187±0.013	0.106±0.007	0.222±0.015	0.310±0.028
GLADE	$(0.3, 1)$	0.160±0.022	0.230±0.007	0.176±0.009	0.097±0.006	0.194±0.013	0.277±0.034
	$(0.3, 2)$	0.179±0.026	0.234±0.009	0.177±0.011	0.100±0.006	0.204±0.012	0.296±0.038
	$(0.3, 3)$	0.177±0.030	0.239±0.006	0.180±0.010	0.103±0.008	0.206±0.016	0.309±0.033
	$(0.7, 1)$	0.190±0.025	0.236±0.007	0.179±0.010	0.102±0.006	0.208±0.013	0.300±0.032
	$(0.7, 2)$	0.205±0.037	0.249±0.009	0.185±0.010	0.107±0.006	0.234±0.017	0.321±0.021
	$(0.7, 3)$	0.227±0.024	0.252±0.008	0.189±0.013	0.113±0.008	0.250±0.016	0.348±0.032
MSPML-TLR	$(0.3, 1)$	0.212±0.038	0.343±0.005	0.245±0.011	0.485±0.004	0.225±0.011	0.222±0.026
	$(0.3, 2)$	0.232±0.037	0.348±0.007	0.245±0.013	0.485±0.004	0.247±0.015	0.235±0.014
	$(0.3, 3)$	0.244±0.030	0.348±0.008	0.250±0.010	0.485±0.004	0.249±0.013	0.264±0.029
	$(0.7, 1)$	0.239±0.033	0.346±0.007	0.247±0.012	0.485±0.004	0.234±0.017	0.228±0.021
	$(0.7, 2)$	0.249±0.031	0.354±0.008	0.255±0.011	0.495±0.027	0.267±0.013	0.258±0.020
	$(0.7, 3)$	0.346±0.034	0.363±0.006	0.261±0.014	0.440±0.110	0.298±0.012	0.289±0.020
VADIS	$(0.3, 1)$	0.160±0.020	0.273±0.008	0.173±0.009	0.095±0.007	0.189±0.012	0.258±0.026
	$(0.3, 2)$	0.179±0.023	0.272±0.008	0.174±0.010	0.098±0.008	0.199±0.012	0.280±0.033
	$(0.3, 3)$	0.175±0.026	0.274±0.008	0.178±0.010	0.098±0.007	0.201±0.017	0.288±0.030
	$(0.7, 1)$	0.187±0.028	0.273±0.007	0.177±0.010	0.098±0.007	0.202±0.013	0.286±0.024
	$(0.7, 2)$	0.207±0.035	0.277±0.010	0.182±0.010	0.101±0.006	0.228±0.016	0.304±0.020
	$(0.7, 3)$	0.222±0.024	0.276±0.007	0.184±0.013	0.106±0.008	0.242±0.015	0.330±0.030
GeoCURVE	$(0.3, 1)$	**0.144±0.021**	**0.179±0.006**	**0.157±0.010**	**0.081±0.007**	**0.143±0.013**	**0.147±0.013**
	$(0.3, 2)$	0.160±0.021	**0.182±0.007**	**0.158±0.011**	**0.084±0.008**	**0.145±0.014**	**0.158±0.014**
	$(0.3, 3)$	0.150±0.024	**0.184±0.006**	**0.160±0.010**	**0.084±0.007**	**0.145±0.014**	**0.159±0.017**
	$(0.7, 1)$	0.157±0.029	**0.183±0.006**	**0.160±0.011**	**0.083±0.007**	**0.146±0.014**	**0.155±0.014**
	$(0.7, 2)$	**0.174±0.027**	**0.190±0.007**	**0.165±0.010**	**0.086±0.006**	**0.156±0.013**	**0.162±0.016**
	$(0.7, 3)$	**0.188±0.029**	**0.192±0.005**	**0.166±0.013**	**0.091±0.008**	**0.164±0.012**	**0.170±0.013**

Overall, these tests provide consistent statistical evidence for the effectiveness of GeoCURVE across multiple evaluation criteria.

D.6 Parameter Sensitivity

We investigate the sensitivity of GeoCURVE to the two capacity parameters γ_r and γ_c by varying one parameter at a time while fixing the other at its selected value. Fig. 4 report the AP results on all six datasets under different ambiguity configurations.

Overall, GeoCURVE exhibits smooth sensitivity to both parameters, although the preferred ranges vary across datasets. Small-to-moderate capacity values are generally favorable, whereas overly

Table 5: Coverage (COV ↓) results under different ambiguity configurations. The best and second-best results in each configuration are shown in bold and underlined, respectively.

Method	(p, r)	Emotions	EspGame5k	Yeast	OBJECT	Human	Plant
FastGRAIL	$(0.3, 1)$	0.307±0.034	0.415±0.010	0.464±0.014	0.169±0.009	0.182±0.007	0.169±0.016
	$(0.3, 2)$	0.315±0.033	0.419±0.010	0.467±0.014	0.172±0.009	0.185±0.010	0.182±0.011
	$(0.3, 3)$	0.314±0.028	0.425±0.009	0.469±0.015	0.174±0.010	0.191±0.010	0.184±0.014
	$(0.7, 1)$	0.317±0.022	0.423±0.009	0.471±0.015	0.173±0.009	0.188±0.009	0.184±0.019
	$(0.7, 2)$	0.320±0.028	0.436±0.012	0.474±0.013	0.178±0.008	0.194±0.009	0.190±0.020
	$(0.7, 3)$	0.327±0.017	0.442±0.013	0.478±0.015	0.185±0.009	0.202±0.013	0.200±0.014
TMvML	$(0.3, 1)$	0.316±0.024	**0.374±0.011**	0.466±0.014	0.213±0.005	0.164±0.009	0.166±0.016
	$(0.3, 2)$	0.330±0.028	0.381±0.012	0.478±0.016	0.215±0.007	0.176±0.005	0.177±0.018
	$(0.3, 3)$	0.334±0.026	**0.384±0.011**	0.480±0.014	0.220±0.008	0.175±0.012	0.176±0.024
	$(0.7, 1)$	0.318±0.022	**0.381±0.006**	0.476±0.014	0.219±0.008	0.174±0.007	0.177±0.021
	$(0.7, 2)$	0.344±0.026	0.415±0.009	0.483±0.017	0.221±0.012	0.182±0.011	0.192±0.019
	$(0.7, 3)$	0.340±0.015	0.407±0.012	0.490±0.021	0.228±0.010	0.195±0.010	0.198±0.016
GRADIS	$(0.3, 1)$	0.288±0.023	0.413±0.010	0.460±0.013	0.182±0.005	0.167±0.011	0.159±0.010
	$(0.3, 2)$	**0.284±0.020**	0.422±0.011	0.461±0.012	0.186±0.007	0.168±0.012	0.173±0.014
	$(0.3, 3)$	**0.290±0.017**	0.430±0.009	0.467±0.015	0.189±0.006	0.169±0.012	0.176±0.016
	$(0.7, 1)$	**0.282±0.023**	0.424±0.009	0.467±0.016	0.190±0.007	0.169±0.011	0.174±0.019
	$(0.7, 2)$	**0.312±0.024**	0.445±0.008	0.483±0.018	0.200±0.008	**0.171±0.009**	0.186±0.015
	$(0.7, 3)$	0.330±0.027	0.457±0.010	0.492±0.016	0.204±0.008	**0.178±0.012**	0.180±0.012
FIMAN	$(0.3, 1)$	0.314±0.026	0.446±0.011	0.481±0.016	0.181±0.008	0.202±0.010	0.252±0.030
	$(0.3, 2)$	0.315±0.026	0.453±0.009	0.486±0.016	0.185±0.009	0.208±0.011	0.267±0.035
	$(0.3, 3)$	0.313±0.026	0.456±0.011	0.484±0.018	0.185±0.010	0.208±0.014	0.267±0.018
	$(0.7, 1)$	0.323±0.021	0.453±0.009	0.489±0.016	0.188±0.008	0.216±0.013	0.280±0.026
	$(0.7, 2)$	0.335±0.021	0.470±0.014	0.489±0.020	0.193±0.008	0.235±0.012	0.289±0.023
	$(0.7, 3)$	**0.321±0.021**	0.471±0.015	0.488±0.017	0.196±0.009	0.236±0.015	0.304±0.028
GLADE	$(0.3, 1)$	0.297±0.024	0.451±0.012	0.468±0.016	0.180±0.009	0.208±0.012	0.272±0.035
	$(0.3, 2)$	0.317±0.032	0.456±0.012	0.471±0.017	0.187±0.009	0.218±0.012	0.290±0.037
	$(0.3, 3)$	0.312±0.033	0.463±0.013	0.474±0.015	0.190±0.011	0.219±0.016	0.302±0.033
	$(0.7, 1)$	0.326±0.027	0.458±0.010	0.476±0.017	0.189±0.008	0.222±0.012	0.293±0.033
	$(0.7, 2)$	0.341±0.039	0.476±0.013	0.487±0.014	0.196±0.007	0.248±0.016	0.311±0.021
	$(0.7, 3)$	0.353±0.026	0.477±0.012	0.490±0.017	0.206±0.009	0.263±0.016	0.340±0.033
MSPML-TLR	$(0.3, 1)$	0.327±0.021	0.557±0.005	0.506±0.011	0.660±0.005	0.236±0.013	0.217±0.025
	$(0.3, 2)$	0.345±0.022	0.565±0.010	0.509±0.008	0.660±0.005	0.259±0.016	0.229±0.012
	$(0.3, 3)$	0.357±0.022	0.567±0.009	0.511±0.006	0.660±0.005	0.261±0.013	0.257±0.026
	$(0.7, 1)$	0.356±0.015	0.566±0.009	0.510±0.012	0.660±0.005	0.245±0.016	0.223±0.020
	$(0.7, 2)$	0.373±0.024	0.577±0.009	0.520±0.014	0.665±0.017	0.278±0.012	0.251±0.019
	$(0.7, 3)$	0.446±0.027	0.590±0.006	0.525±0.010	0.601±0.128	0.305±0.010	0.283±0.017
VADIS	$(0.3, 1)$	0.298±0.021	0.503±0.008	0.464±0.016	0.177±0.010	0.203±0.012	0.254±0.027
	$(0.3, 2)$	0.318±0.030	0.502±0.008	0.466±0.016	0.182±0.011	0.213±0.012	0.275±0.033
	$(0.3, 3)$	0.310±0.029	0.506±0.010	0.471±0.015	0.183±0.010	0.215±0.016	0.282±0.029
	$(0.7, 1)$	0.324±0.028	0.503±0.008	0.471±0.018	0.183±0.009	0.216±0.013	0.280±0.025
	$(0.7, 2)$	0.342±0.037	0.508±0.009	0.480±0.012	0.188±0.009	0.243±0.015	0.296±0.020
	$(0.7, 3)$	0.349±0.027	0.509±0.009	0.483±0.017	0.195±0.010	0.255±0.015	0.322±0.031
GeoCURVE	$(0.3, 1)$	**0.286±0.016**	0.375±0.009	**0.442±0.012**	**0.159±0.010**	**0.158±0.012**	**0.148±0.014**
	$(0.3, 2)$	0.300±0.024	**0.378±0.012**	**0.444±0.015**	**0.163±0.012**	**0.160±0.015**	**0.158±0.014**
	$(0.3, 3)$	**0.290±0.022**	**0.384±0.010**	**0.446±0.015**	**0.162±0.011**	**0.161±0.015**	**0.159±0.017**
	$(0.7, 1)$	0.294±0.025	**0.381±0.011**	**0.448±0.018**	**0.162±0.009**	**0.161±0.014**	**0.155±0.015**
	$(0.7, 2)$	0.314±0.027	**0.397±0.010**	**0.455±0.011**	**0.167±0.009**	0.172±0.012	**0.162±0.016**
	$(0.7, 3)$	0.323±0.026	**0.396±0.013**	**0.456±0.018**	**0.174±0.010**	0.180±0.013	**0.170±0.013**

large values may gradually reduce AP, particularly under stronger ambiguity. No sharply isolated optimum is observed over the evaluated ranges.

D.7 Additional Analysis of Candidate Reliability and Correction

We provide additional details for the analysis of candidate reliability and correction, together with results on the five datasets omitted from the main text. All experiments are conducted after model fitting. For each dataset and ambiguity setting, the experiments are computed over the ten outer folds.

Table 6: Hamming loss (HL ↓) results under different ambiguity configurations. The best and second-best results in each configuration are shown in bold and underlined, respectively.

Method	(p, r)	Emotions	EspGame5k	Yeast	OBJECT	Human	Plant
FastGRAIL	$(0.3, 1)$	0.224±0.017	0.061±0.001	0.222±0.006	0.052±0.002	0.107±0.003	0.105±0.007
	$(0.3, 2)$	0.231±0.020	0.061±0.001	0.222±0.007	0.053±0.002	0.108±0.001	0.107±0.007
	$(0.3, 3)$	0.217±0.016	0.061±0.001	0.222±0.007	0.053±0.001	0.108±0.004	0.112±0.009
	$(0.7, 1)$	0.227±0.013	0.061±0.001	0.223±0.006	0.052±0.002	0.108±0.003	0.109±0.006
	$(0.7, 2)$	0.226±0.014	0.062±0.001	0.223±0.006	0.053±0.001	0.109±0.003	0.112±0.004
	$(0.7, 3)$	0.245±0.023	0.063±0.001	0.222±0.006	0.053±0.002	0.110±0.004	0.121±0.004
TMvML	$(0.3, 1)$	0.236±0.013	0.066±0.001	0.216±0.008	0.065±0.002	0.124±0.003	0.125±0.004
	$(0.3, 2)$	0.241±0.012	0.067±0.001	0.214±0.007	0.066±0.002	0.125±0.003	0.129±0.006
	$(0.3, 3)$	0.241±0.013	0.067±0.001	0.215±0.009	0.066±0.002	0.127±0.003	0.134±0.008
	$(0.7, 1)$	0.249±0.009	0.067±0.001	0.215±0.007	0.066±0.002	0.127±0.003	0.134±0.011
	$(0.7, 2)$	0.245±0.015	0.072±0.001	0.212±0.009	0.068±0.002	0.134±0.003	0.152±0.007
	$(0.7, 3)$	0.256±0.022	0.072±0.001	0.216±0.008	0.070±0.002	0.141±0.005	0.160±0.006
GRADIS	$(0.3, 1)$	**0.186±0.016**	**0.052±0.001**	**0.194±0.008**	**0.043±0.001**	**0.079±0.003**	**0.075±0.002**
	$(0.3, 2)$	**0.179±0.014**	**0.051±0.001**	**0.196±0.007**	**0.043±0.001**	**0.080±0.003**	**0.075±0.003**
	$(0.3, 3)$	**0.194±0.025**	**0.051±0.001**	**0.195±0.007**	**0.043±0.001**	**0.080±0.002**	**0.074±0.003**
	$(0.7, 1)$	**0.192±0.019**	**0.052±0.001**	**0.194±0.007**	**0.043±0.001**	**0.080±0.002**	**0.075±0.004**
	$(0.7, 2)$	0.236±0.022	**0.052±0.001**	**0.195±0.007**	**0.044±0.001**	**0.079±0.003**	**0.074±0.003**
	$(0.7, 3)$	0.403±0.031	**0.052±0.001**	**0.202±0.006**	**0.044±0.001**	**0.081±0.003**	**0.076±0.004**
FIMAN	$(0.3, 1)$	0.228±0.019	0.111±0.002	0.211±0.008	0.075±0.003	0.158±0.006	0.219±0.014
	$(0.3, 2)$	0.227±0.017	0.116±0.003	0.213±0.008	0.076±0.003	0.164±0.005	0.234±0.012
	$(0.3, 3)$	0.233±0.019	0.118±0.003	0.209±0.007	0.076±0.003	0.167±0.006	0.230±0.015
	$(0.7, 1)$	0.235±0.016	0.119±0.003	0.211±0.007	0.076±0.003	0.176±0.008	0.263±0.007
	$(0.7, 2)$	0.238±0.011	0.130±0.004	0.211±0.008	0.078±0.003	0.200±0.007	0.280±0.018
	$(0.7, 3)$	0.248±0.015	0.141±0.004	0.211±0.008	0.081±0.004	0.210±0.005	0.281±0.018
GLADE	$(0.3, 1)$	0.279±0.025	0.268±0.007	0.248±0.010	0.145±0.006	0.260±0.007	0.388±0.012
	$(0.3, 2)$	0.284±0.025	0.283±0.006	0.250±0.011	0.152±0.006	0.296±0.009	0.431±0.018
	$(0.3, 3)$	0.290±0.034	0.296±0.007	0.251±0.011	0.161±0.007	0.316±0.011	0.438±0.015
	$(0.7, 1)$	0.307±0.030	0.288±0.009	0.254±0.011	0.157±0.007	0.305±0.008	0.448±0.009
	$(0.7, 2)$	0.323±0.023	0.322±0.009	0.259±0.010	0.180±0.007	0.367±0.010	0.472±0.013
	$(0.7, 3)$	0.338±0.014	0.347±0.007	0.276±0.011	0.199±0.007	0.405±0.009	0.486±0.023
MSPML-TLR	$(0.3, 1)$	0.677±0.012	0.418±0.013	0.652±0.007	0.066±0.000	0.766±0.015	0.770±0.026
	$(0.3, 2)$	0.679±0.014	0.448±0.011	0.655±0.005	0.066±0.000	0.819±0.013	0.836±0.026
	$(0.3, 3)$	0.678±0.013	0.468±0.011	0.657±0.007	0.066±0.000	0.833±0.010	0.863±0.022
	$(0.7, 1)$	0.680±0.013	0.451±0.010	0.654±0.004	0.066±0.000	0.833±0.013	0.846±0.019
	$(0.7, 2)$	0.678±0.012	0.494±0.015	0.654±0.007	0.132±0.008	0.843±0.014	0.890±0.018
	$(0.7, 3)$	0.677±0.012	0.514±0.012	0.654±0.006	0.258±0.294	0.850±0.014	0.884±0.011
VADIS	$(0.3, 1)$	0.208±0.019	0.098±0.004	0.209±0.009	0.058±0.002	0.126±0.005	0.159±0.013
	$(0.3, 2)$	0.220±0.023	0.098±0.004	0.210±0.009	0.058±0.002	0.131±0.004	0.174±0.011
	$(0.3, 3)$	0.227±0.017	0.099±0.005	0.209±0.008	0.059±0.002	0.136±0.006	0.181±0.013
	$(0.7, 1)$	0.230±0.024	0.099±0.004	0.211±0.009	0.058±0.002	0.134±0.005	0.179±0.007
	$(0.7, 2)$	0.240±0.024	0.100±0.005	0.211±0.007	0.059±0.002	0.151±0.005	0.196±0.008
	$(0.7, 3)$	0.378±0.065	0.101±0.005	0.214±0.008	0.061±0.003	0.166±0.003	0.212±0.016
GeoCURVE	$(0.3, 1)$	0.193±0.021	0.060±0.001	0.200±0.006	0.049±0.002	0.106±0.004	0.113±0.007
	$(0.3, 2)$	0.198±0.026	0.060±0.001	0.202±0.007	0.050±0.002	0.106±0.004	0.113±0.006
	$(0.3, 3)$	0.200±0.018	0.060±0.001	0.201±0.006	0.050±0.002	0.107±0.005	0.115±0.008
	$(0.7, 1)$	0.199±0.020	0.060±0.001	0.202±0.007	0.050±0.002	0.107±0.004	0.114±0.008
	$(0.7, 2)$	**0.204±0.023**	0.060±0.001	0.203±0.005	0.050±0.002	0.107±0.003	0.115±0.007
	$(0.7, 3)$	**0.228±0.018**	0.061±0.001	0.204±0.007	0.051±0.002	0.109±0.004	0.119±0.007

Reliability–relevance relation. For each candidate entry with $Y_{ij} = 1$, GeoCURVE produces a reliability score $\hat{\eta}_{ij}$. Candidate entries are grouped according to their reliability scores, and the empirical relevance rate within each interval is computed as the fraction satisfying $Y_{ij}^{\text{true}} = 1$. The interval statistics are aggregated across folds according to the number of candidate entries they contain. This experiment examines whether the estimated reliability meaningfully orders candidate labels according to their latent relevance, rather than whether it constitutes a calibrated probability.

Negative-like mixture assumption. Using the clean labels only for diagnosis, we define the false-candidate, true-candidate, and trusted-negative sets as

$$\mathcal{F} = \{(i, j) : Y_{ij} = 1,\ Y_{ij}^{\text{true}} = 0\}, \tag{98}$$

Table 7: One-error (OE ↓) results under different ambiguity configurations. The best and second-best results in each configuration are shown in bold and underlined, respectively.

Method	(p, r)	Emotions	EspGame5k	Yeast	OBJECT	Human	Plant
FastGRAIL	$(0.3, 1)$	0.307±0.039	0.589±0.015	0.229±0.020	0.305±0.022	0.590±0.021	0.620±0.037
	$(0.3, 2)$	0.307±0.050	0.589±0.019	0.228±0.023	0.306±0.021	0.587±0.023	0.611±0.044
	$(0.3, 3)$	0.303±0.038	0.596±0.018	0.230±0.021	0.302±0.021	0.590±0.027	0.627±0.048
	$(0.7, 1)$	0.317±0.055	0.595±0.015	0.229±0.019	0.303±0.021	0.585±0.021	0.630±0.025
	$(0.7, 2)$	0.300±0.049	0.604±0.017	0.227±0.020	0.306±0.023	0.587±0.031	0.627±0.035
	$(0.7, 3)$	0.329±0.074	0.609±0.022	0.235±0.016	0.308±0.020	0.592±0.022	0.645±0.021
TMvML	$(0.3, 1)$	0.322±0.062	0.581±0.020	0.249±0.021	0.379±0.025	0.568±0.020	0.603±0.027
	$(0.3, 2)$	0.334±0.046	0.587±0.019	0.246±0.026	0.391±0.015	0.567±0.017	0.605±0.024
	$(0.3, 3)$	0.334±0.066	0.584±0.020	0.247±0.024	0.386±0.023	0.565±0.012	0.621±0.035
	$(0.7, 1)$	0.312±0.043	0.586±0.016	0.245±0.023	0.387±0.020	0.573±0.016	0.608±0.038
	$(0.7, 2)$	0.332±0.057	0.569±0.025	0.245±0.027	0.387±0.024	0.581±0.020	0.632±0.039
	$(0.7, 3)$	0.361±0.035	0.564±0.020	0.252±0.023	0.382±0.027	0.577±0.018	0.639±0.038
GRADIS	$(0.3, 1)$	0.247±0.064	0.555±0.015	**0.216±0.017**	0.282±0.021	0.554±0.017	0.625±0.032
	$(0.3, 2)$	0.245±0.058	0.556±0.018	0.216±0.021	0.279±0.018	0.550±0.028	0.635±0.048
	$(0.3, 3)$	**0.243±0.068**	0.564±0.018	0.223±0.022	0.283±0.016	0.557±0.026	0.639±0.059
	$(0.7, 1)$	**0.248±0.056**	0.560±0.012	0.219±0.020	0.283±0.019	0.556±0.018	0.636±0.039
	$(0.7, 2)$	0.293±0.060	0.559±0.023	0.217±0.017	0.291±0.020	0.567±0.017	0.633±0.037
	$(0.7, 3)$	0.311±0.046	0.578±0.014	0.225±0.025	0.294±0.020	0.577±0.029	0.641±0.052
FIMAN	$(0.3, 1)$	0.271±0.060	0.560±0.023	**0.216±0.025**	0.304±0.026	0.574±0.019	0.686±0.023
	$(0.3, 2)$	0.277±0.052	0.559±0.021	0.220±0.008	0.300±0.027	0.575±0.020	0.683±0.036
	$(0.3, 3)$	0.283±0.039	0.567±0.018	0.220±0.027	0.305±0.026	0.586±0.020	0.704±0.034
	$(0.7, 1)$	0.284±0.065	0.568±0.017	0.218±0.024	0.304±0.021	0.584±0.014	0.708±0.035
	$(0.7, 2)$	0.295±0.038	0.567±0.017	0.223±0.025	0.309±0.026	0.601±0.023	0.724±0.026
	$(0.7, 3)$	0.321±0.048	0.584±0.017	**0.222±0.022**	0.307±0.020	0.610±0.029	0.740±0.038
GLADE	$(0.3, 1)$	0.270±0.070	0.572±0.017	0.226±0.025	0.304±0.025	0.584±0.020	0.706±0.032
	$(0.3, 2)$	0.271±0.065	0.573±0.014	0.228±0.025	0.303±0.028	0.588±0.025	0.701±0.048
	$(0.3, 3)$	0.287±0.043	0.580±0.016	0.230±0.030	0.312±0.028	0.597±0.032	0.725±0.040
	$(0.7, 1)$	0.295±0.060	0.578±0.017	0.226±0.030	0.306±0.024	0.594±0.022	0.732±0.041
	$(0.7, 2)$	0.312±0.064	0.584±0.019	0.228±0.026	0.313±0.028	0.620±0.029	0.767±0.032
	$(0.7, 3)$	0.383±0.066	0.604±0.026	0.237±0.021	0.316±0.024	0.640±0.025	0.783±0.056
MSPML-TLR	$(0.3, 1)$	0.352±0.076	0.720±0.018	0.318±0.028	0.981±0.007	0.666±0.027	0.707±0.044
	$(0.3, 2)$	0.369±0.056	0.727±0.017	0.318±0.025	0.981±0.007	0.656±0.028	0.701±0.037
	$(0.3, 3)$	0.374±0.054	0.732±0.014	0.326±0.023	0.981±0.007	0.677±0.025	0.709±0.042
	$(0.7, 1)$	0.381±0.068	0.730±0.017	0.316±0.023	0.981±0.007	0.674±0.020	0.695±0.042
	$(0.7, 2)$	0.366±0.071	0.746±0.016	0.325±0.019	0.981±0.008	0.723±0.026	0.710±0.041
	$(0.7, 3)$	0.484±0.066	0.772±0.014	0.328±0.026	0.900±0.159	0.779±0.015	0.722±0.041
VADIS	$(0.3, 1)$	0.266±0.068	0.654±0.027	0.225±0.024	0.303±0.024	0.580±0.017	0.688±0.024
	$(0.3, 2)$	0.278±0.065	0.659±0.024	0.229±0.027	0.302±0.019	0.581±0.026	0.686±0.048
	$(0.3, 3)$	0.283±0.026	0.661±0.023	0.229±0.029	0.306±0.020	0.592±0.030	0.709±0.045
	$(0.7, 1)$	0.285±0.067	0.661±0.022	0.224±0.029	0.304±0.022	0.590±0.025	0.713±0.038
	$(0.7, 2)$	0.308±0.065	0.667±0.020	0.229±0.027	0.307±0.021	0.611±0.024	0.736±0.037
	$(0.7, 3)$	**0.266±0.014**	0.664±0.024	0.230±0.029	0.310±0.023	0.630±0.020	0.767±0.042
GeoCURVE	$(0.3, 1)$	**0.238±0.059**	**0.493±0.023**	**0.216±0.025**	**0.254±0.021**	**0.516±0.025**	**0.596±0.028**
	$(0.3, 2)$	**0.243±0.068**	**0.495±0.018**	**0.211±0.022**	**0.258±0.020**	**0.524±0.021**	**0.588±0.022**
	$(0.3, 3)$	0.250±0.065	**0.495±0.016**	**0.209±0.019**	**0.256±0.020**	**0.527±0.025**	**0.594±0.044**
	$(0.7, 1)$	0.256±0.044	**0.490±0.018**	**0.217±0.028**	**0.256±0.020**	**0.524±0.021**	**0.599±0.040**
	$(0.7, 2)$	**0.273±0.069**	**0.499±0.021**	**0.211±0.025**	**0.257±0.019**	**0.533±0.025**	**0.611±0.025**
	$(0.7, 3)$	0.308±0.058	**0.500±0.022**	0.227±0.024	**0.264±0.024**	**0.533±0.027**	**0.619±0.038**

$$\mathcal{T} = \{(i, j) : Y_{ij} = 1,\ Y_{ij}^{\text{true}} = 1\}, \tag{99}$$

and

$$\mathcal{N} = \{(i, j) : Y_{ij} = 0\}, \tag{100}$$

respectively. We compare their normalized structural-evidence distributions using the Jensen–Shannon (JS) divergence Lin (1991). A smaller JS divergence indicates greater similarity between two distributions. Therefore,

$$D_{\text{JS}}(\mathcal{F}, \mathcal{N}) < D_{\text{JS}}(\mathcal{T}, \mathcal{N}) \tag{101}$$

indicates that false candidates exhibit more negative-like structural evidence than true candidates, providing empirical evidence consistent with the negative-like assumption underlying the reliability estimator.

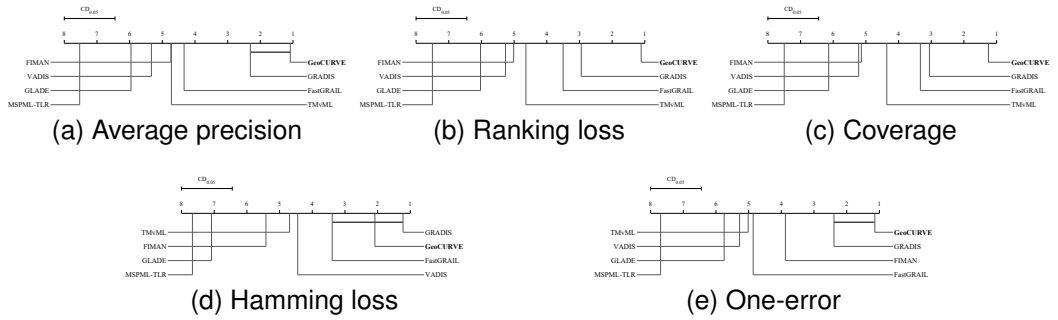


Figure 3: Bonferroni–Dunn comparisons with GeoCURVE as the control on five evaluation metrics. Methods not connected by the thick bar differ significantly from GeoCURVE at $\alpha = 0.05$.

Table 8: Iman–Davenport statistics F_F and the critical value at $\alpha = 0.05$ ($N = 36$, $K = 8$).

Evaluation metric	F_F	Critical value
Average Precision	74.7894	2.0471
Hamming Loss	187.7689	2.0471
One-Error	77.3070	2.0471
Ranking Loss	63.6930	2.0471
Coverage	60.7347	2.0471

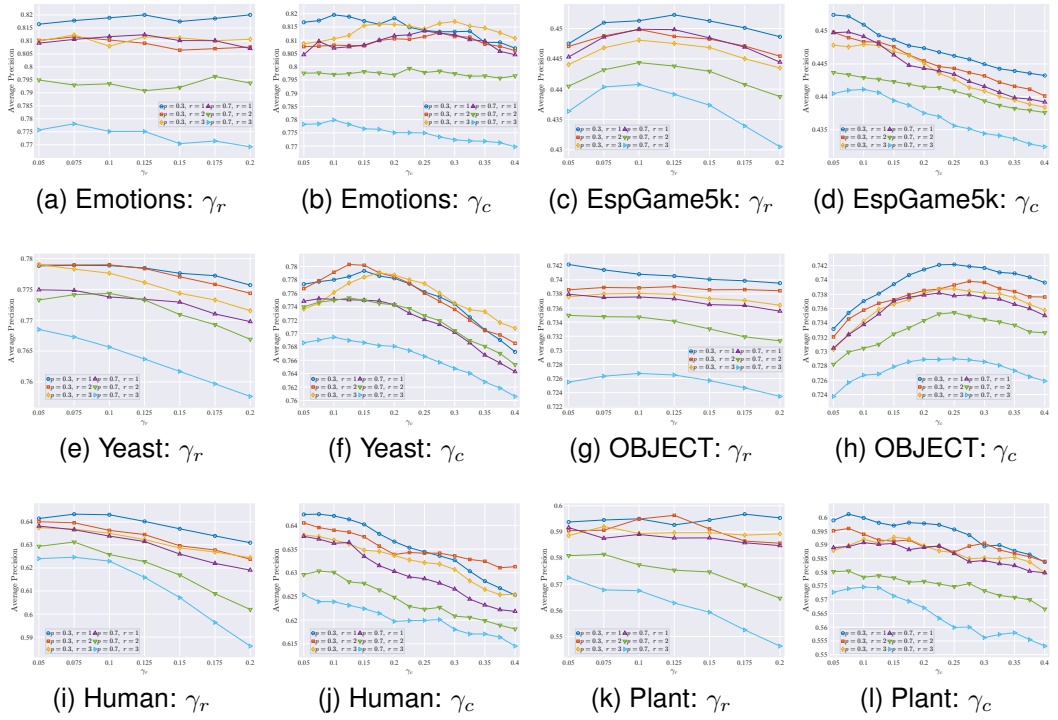


Figure 4: Parameter sensitivity of GeoCURVE with respect to γ_r and γ_c on all six benchmark datasets under different ambiguity configurations.

Correction alignment. We further examine whether reliability estimation improves the direction of correction from the structural reference. The oracle, raw candidate, and reliability-informed corrections are respectively defined as

$$\mathbf{R}_{\text{oracle}} = \mathbf{Y}^{\text{true}} - \mathbf{P}_0, \tag{102}$$

$$\mathbf{R}_Y = \mathbf{Y} - \mathbf{P}_0, \tag{103}$$

and

$$\mathbf{R}_{\mathrm{cs}} = \widetilde{\mathbf{Y}} - \mathbf{P}_0. \tag{104}$$

Their directional agreement with the oracle correction is measured by

$$\mathrm{Align}(\mathbf{R}, \mathbf{R}_{\mathrm{oracle}}) = \frac{\langle \mathbf{R}, \mathbf{R}_{\mathrm{oracle}} \rangle_F}{\|\mathbf{R}\|_F \|\mathbf{R}_{\mathrm{oracle}}\|_F}. \tag{105}$$

Since cosine alignment is invariant to positive rescaling, it isolates the correction direction from its magnitude. For each ambiguity setting, the bars report the mean over the ten outer folds, while the error bars denote the corresponding standard deviation.

Figure 5 reports the results on OBJECT, EspGame5k, Yeast, Human, and Plant. The same overall behavior observed on Emotions is consistently reproduced. Higher reliability scores are generally associated with higher empirical relevance rates; false candidates remain substantially closer to trusted negatives than true candidates in structural-evidence space; and reliability-informed corrections exhibit stronger oracle alignment than raw candidate corrections. Across all six datasets and four ambiguity settings, the negative-like divergence relation holds in all 240 outer folds, while the reliability-informed correction achieves higher oracle alignment in 223 of 240 folds. These results further support the role of candidate reliability in determining an appropriate correction direction before it is subsequently shaped by the shared multi-view geometry.

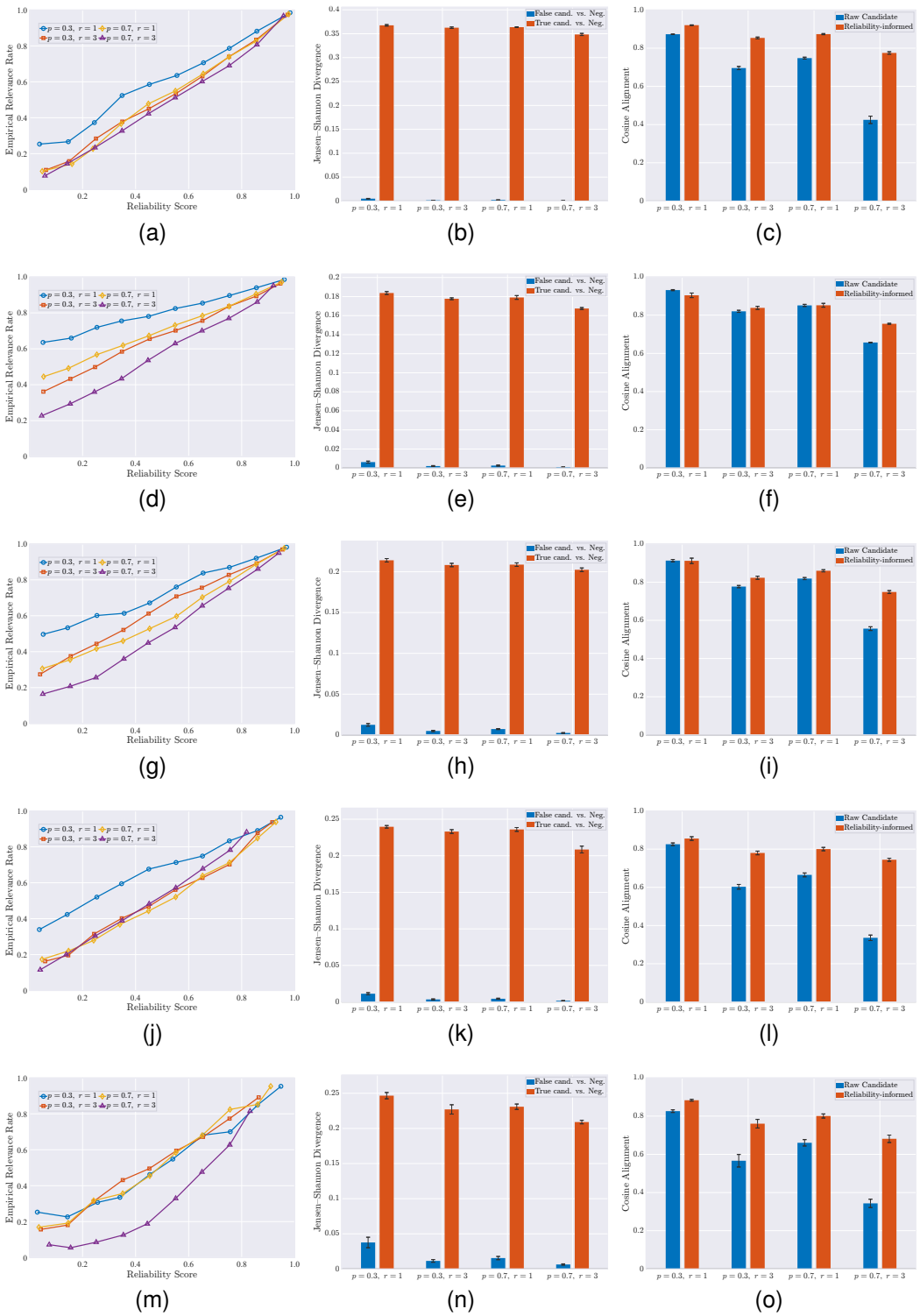


Figure 5: Additional analysis of candidate reliability and correction on OBJECT, EspGame5k, Yeast, Human, and Plant from top to bottom. From left to right, the columns show the reliability–relevance relation, the Jensen–Shannon divergence for the negative-like mixture assumption, and the oracle correction alignment.