# ANTECEDENT-ALIGNED RULE TRANSFER WITH PARTIAL-OVERLAP CONSEQUENT FUSION FOR MULTI-CENTER TSK FUZZY CLASSIFICATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Multi-center Takagi–Sugeno–Kang (TSK) fuzzy classification faces challenges arising from asymmetric sample sizes and partially overlapping local rule bases, as independently learned antecedents may not denote comparable fuzzy regions across centers. We propose MCRT-TSK, which establishes shared antecedent identities for semantic rule matching while allowing each center to retain only locally supported rules. At the target center, a sample-count-adaptive criterion combines source-side support with target activation evidence, and consequent blocks are fused only from sources containing the matched antecedent. A rule-level prior and a response-level anchor regularize a convex target consequent-estimation problem with a closed-form solution. On 18 UCI datasets, MCRT-TSK achieves the highest or tied-highest target-test ACC among methods with valid outputs on 16 of the 18 datasets. Across three stress-test families, shared antecedent transfer yields positive mean ACC, Macro-F1, and G-mean gains over local-target training; identity-preserving consequent transfer also yields positive mean margins over a within-order shuffled-prior control in all nine task-family–metric summaries. Statistical, ablation, efficiency, and rule-level evidence supports the full MCRT-TSK under the simulated-center protocol and aligned rule transfer in the stress tests.

## 1 INTRODUCTION

Multi-center supervised classification is common in data-sensitive scientific and medical applications, where center roles, sample sizes, and acquisition processes differ. Deployments may exhibit covariate, label-prior, or concept shift. Our primary benchmark isolates a TSK-specific challenge—structural mismatch among locally retained fuzzy rules under asymmetric allocation—while complementary stress tests cover controlled feature and label heterogeneity and natural Heart sites. In TSK classifiers (Takagi & Sugeno, 1985; Sugeno & Kang, 1988), reliable transfer requires matched antecedent semantics and sufficient local activation support; otherwise, corresponding rules may represent different input regions. A global rule base fixes semantics but retains locally unsupported rules; independent rule bases preserve local specificity but obscure correspondence. Partial overlap calls for aligned antecedent identities, support-aware selection, and consequent fusion restricted to matched rules.

Federated and multi-source methods generally aggregate parameters or updates without checking antecedent correspondence (Guo et al., 2024), while domain adaptation primarily aligns features or representations rather than TSK antecedent–consequent structures (Li et al., 2024b; Zhao et al., 2024). Recent fuzzy and TSK methods improve adaptation, interpretability, alignment, or consequent learning (Ma et al., 2024; Tian et al., 2025; Zhang et al., 2024; Li et al., 2025; Wang et al., 2024; 2026; Yao et al., 2026), but do not jointly address explicit antecedent identities, support-aware selection under uneven support and asymmetric target allocation, and consequent fusion under partial rule overlap.

We propose a Multi-Center Rule-Transfer TSK Fuzzy System (MCRT-TSK). The main center defines shared semantics and bounded-order antecedent identities: feature subsets with fuzzy-set index assignments. Centers retain only locally supported identities, which serve as matching keys for



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.

transferring antecedent–consequent rule blocks. Target selection combines source support and target activation, increasingly favoring target evidence as its sample size grows.

Only sources retaining each target identity contribute consequents, weighted by support-profile compatibility with the target-selected structure. A response-level anchor stabilizes learning with limited target data. With fixed antecedents and target rules, consequent optimization is convex quadratic with a closed-form ridge-type solution. Our contributions are: (i) We formulate cross-center TSK transfer through shared antecedent semantics and explicit identities, enabling rule comparison without identical local rule bases. (ii) We propose a support-aware, sample-count-adaptive selection strategy that balances a source-support proxy with target activation under asymmetric allocation. (iii) We develop a partial-overlap fusion scheme that aggregates consequent blocks only from sources sharing the target identity while preserving closed-form target optimization.

## 2 Related Work

TSK fuzzy systems model nonlinear predictions via IF–THEN rules, each pairing a fuzzy antecedent with a consequent (e.g., IF $x_1$ is $A_{k1}$ and $\cdots$ and $x_d$ is $A_{kd}$, THEN $\mathbf{g}_k(\mathbf{x})$), learned from data or neuro-fuzzy architectures (Wang & Mendel, 1992; Jang, 1993). While this explicit structure aids interpretability, cross-center consequent reuse requires aligned antecedent semantics and local rule support. Collaborative transfer methods do not enforce such rule-wise correspondence. Federated learning and federated transfer learning reduce raw-data sharing, but aggregate parameters or updates without matching antecedent-defined fuzzy regions (McMahan et al., 2017; Guo et al., 2024). Source-free and multi-source domain adaptation reduce distributional discrepancies, yet align features or predictions rather than matched TSK rule blocks (Pan & Yang, 2010; Li et al., 2024b; Zhao et al., 2024).

TSK-based transfer methods adapt fuzzy representations, calibrate multi-center knowledge, or couple distributional and label-space structure (Xu et al., 2021; Wang et al., 2024; 2026). The most directly related comparator, MTSK-DA, uses discriminative alignment and attention-guided graph regularization to promote class separation and cross-center relation modeling (Yao et al., 2026). Its transfer interface is distributional and graph-based rather than an explicit mapping between antecedent identities and consequent blocks. MCRT-TSK instead matches shared antecedent identities, selects them using local support, and fuses consequents only across sources retaining the same rule. The supplementary material provides a detailed method-level comparison.

## 3 The Proposed MCRT-TSK Method

### 3.1 Problem Formulation

We consider an asymmetric collaborative setting with a main center, multiple auxiliary centers, and a target center assigned the smallest sample share. These roles are simulated in the primary benchmark and instantiated by the four site-specific Heart cohorts in the natural-center stress test. MCRT-TSK separates shared antecedent semantics, locally retained identities, and consequent transfer across matched antecedents. Figure 1 summarizes the model architecture.

For notational compactness, the superscript $*$ denotes a generic center index, where $* \in \{m, (1), \ldots, (A), t\}$. Here, $m$, $(a)$, and $t$ represent the main center, the $a$-th auxiliary center, and the target center, respectively. All centers share feature definitions, scaling, and the $C$-class label space. For any center $*$, let $\mathcal{D}^* = \{(\mathbf{x}_i^*, y_i^*)\}_{i=1}^{N_*}$, where $\mathbf{x}_i^* \in [0,1]^d$ is a normalized $d$-dimensional feature vector, $N_*$ is the sample size, and $y_i^* \in \{1, \ldots, C\}$ is the class label. The one-hot labels are collected in $\mathbf{Y}^* \in \{0,1\}^{N_* \times C}$.

### 3.2 Main-Center Rule Construction

**Feature-Level Adaptive Fuzzy Partitioning.** For the $j$-th feature, the main-center fuzzy partition is defined as $\Psi_j^m = \{(c_{j,l}^m, \sigma_{j,l}^m)\}_{l=1}^{L_j}$, where $L_j$ is the number of fuzzy sets and $c_{j,l}^m$ and $\sigma_{j,l}^m$ denote the center and width of the $l$-th fuzzy set. Candidate granularities are restricted to $\mathcal{L} = \{1, 2, \ldots, L_{\max}\}$. For each $L \in \mathcal{L}$, let $\mathcal{C}_{j,L}$ be the ordered center set returned by the deterministic one-dimensional

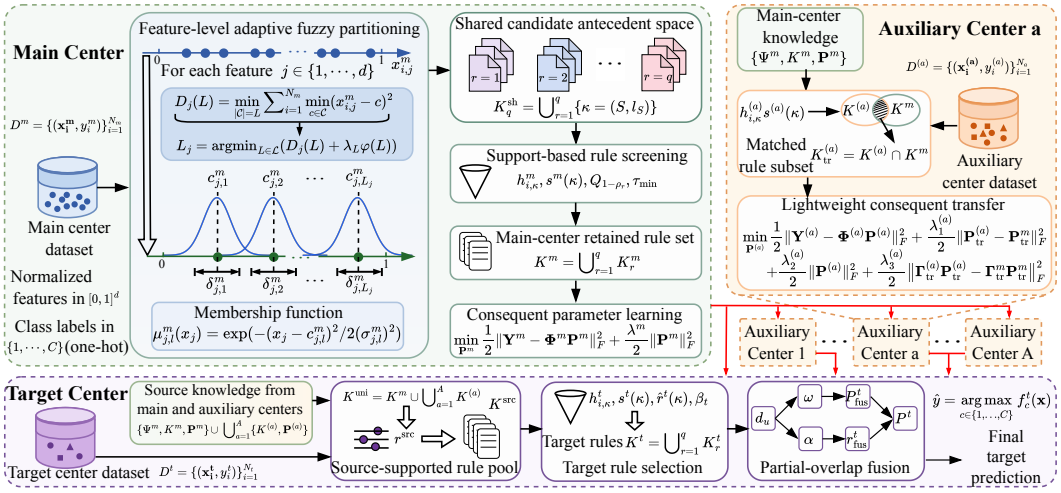


Figure 1: Overview of the proposed MCRT-TSK method.

Lloyd routine used in the experiments. Its achieved main-center distortion is

$$D_j(L) = \sum_{i=1}^{N_m} \min_{c\in\mathcal{C}_{j,L}} (x_{i,j}^m - c)^2, \tag{1}$$

which does not guarantee globally optimal one-dimensional quantization. A nondecreasing complexity penalty $\varphi(L)$ limits granularity:

$$L_j = \arg\min_{L\in\mathcal{L}}(D_j(L) + \lambda_L \varphi(L)), \tag{2}$$

where $\lambda_L \geq 0$ controls the penalty and we use $\varphi(L) = L$. For ordered centers $c_{j,1}^m \leq \cdots \leq c_{j,L_j}^m$, local spacings determine widths via $\sigma_{j,l}^m = \eta \max\{\delta_{j,l}^m, \varepsilon\}$ and

$$\delta_{j,l}^m = \begin{cases} \max\{\max_i x_{i,j}^m - \min_i x_{i,j}^m, \varepsilon\}, & L_j = 1, \\ c_{j,2}^m - c_{j,1}^m, & L_j > 1,\ l = 1, \\ \min\{c_{j,l}^m - c_{j,l-1}^m,\ c_{j,l+1}^m - c_{j,l}^m\}, & 1 < l < L_j, \\ c_{j,L_j}^m - c_{j,L_j-1}^m, & L_j > 1,\ l = L_j, \end{cases} \tag{3}$$

where $\eta > 0$ controls the overlap of adjacent fuzzy sets and $\varepsilon$ prevents degenerate zero-width Gaussian sets. The Gaussian membership function of the $l$-th fuzzy set on the $j$-th feature is $\mu_{j,l}^m(x_j) = \exp(-(x_j - c_{j,l}^m)^2/(2(\sigma_{j,l}^m)^2))$. The collection $\Psi^m = \{\Psi_j^m\}_{j=1}^d$ defines the shared antecedent semantics used for cross-center antecedent-identity encoding.

**Shared Candidate Antecedent Space and Main-Center Support-Based Rule Screening.** After feature-level partitioning, enumerating every fuzzy-set combination as a final rule is infeasible because of combinatorial growth. MCRT-TSK therefore constructs a shared candidate antecedent-identity space induced by $\Psi^m$. Let $q$, with $1 \leq q \leq d$, be the maximum rule order. Then $K_q^{\mathrm{sh}} \equiv K_q(\Psi^m) = \bigcup_{r=1}^{q} \{\kappa = (S, l_S) \,|\, S \subseteq \{1,\ldots,d\}, |S| = r, l_j \in \{1,\ldots,L_j\},\ j \in S\}$. Each key $\kappa = (S, l_S)$ defines a shared antecedent identity used by all centers. For any candidate rule $\kappa = (S, l_S) \in K_q^{\mathrm{sh}}$, its unnormalized firing strength on the main-center sample $\mathbf{x}_i^m$ and the corresponding main-center support are defined as

$$h_{i,\kappa}^m = \prod_{j\in S} \mu_{j,l_j}^m(x_{i,j}^m),\ \kappa = (S, l_S), \qquad s^m(\kappa) = \frac{1}{N_m} \sum_{i=1}^{N_m} h_{i,\kappa}^m. \tag{4}$$

The mean support $s^m(\kappa) \in [0,1]$ is sample-size normalized. For main- and auxiliary-center support-based rule screening, we use a hierarchical screening operator $\mathsf{HS}(\mathcal{K}, \zeta)$ with $0 < \rho_1 \leq \cdots \leq \rho_q \leq 1$ and $\tau_{\min} > 0$. Given a rule-score function $\zeta(\kappa)$, for each order $o$, rules are retained when their scores are at least $\max\{Q_{1-\rho_o}, \tau_{\min}\}$, where $Q_{1-\rho_o}$ is the empirical quantile over the nonempty recursively eligible candidates at that order and ties are retained; an $o$-order rule with $o \geq 2$ is considered only when all its $(o-1)$-order subrules have been retained. The main-center locally

retained rule set is $K^m = \mathsf{HS}(K_q^{\mathrm{sh}}, s^m)$. Empty-set fallback and stepwise definitions are in the supplementary material.

**Consequent Learning on the Main-Center Retained Rule Set.** With $K^m$ fixed, the design matrix is constructed from the retained rules. For any $\kappa \in K^m$, the normalized firing strength on the main center is $\tilde{h}_{i,\kappa}^m = h_{i,\kappa}^m / \sum_{\kappa' \in K^m} h_{i,\kappa'}^m$. All normalized-firing denominators are lower-bounded by a fixed $\epsilon_h > 0$ to prevent numerical underflow.

For the zero-order TSK consequent, the rule-level feature is $\phi_{i,\kappa}^m = \tilde{h}_{i,\kappa}^m \in \mathbb{R}$. Stacking the consequent blocks $\mathbf{p}_\kappa^m \in \mathbb{R}^C$ in a fixed rule order yields $\mathbf{P}^m \in \mathbb{R}^{|K^m| \times C}$. Concatenating $\{\phi_{i,\kappa}^m\}$ produces the main-center design matrix $\mathbf{\Phi}^m = \left[\phi_{i,\kappa}^m\right]_{i=1,\ldots,N_m;\, \kappa \in K^m} \in \mathbb{R}^{N_m \times |K^m|}$. The consequent parameters are estimated by ridge regression, with the corresponding closed-form solution:

$$\min_{\mathbf{P}^m} \frac{1}{2}\|\mathbf{Y}^m - \mathbf{\Phi}^m \mathbf{P}^m\|_F^2 + \frac{\lambda^m}{2}\|\mathbf{P}^m\|_F^2, \qquad \mathbf{P}^m = ((\mathbf{\Phi}^m)^T \mathbf{\Phi}^m + \lambda^m \mathbf{I})^{-1} (\mathbf{\Phi}^m)^T \mathbf{Y}^m, \tag{5}$$

where $\mathbf{I} \in \mathbb{R}^{|K^m| \times |K^m|}$ and $\lambda^m > 0$.

### 3.3 Auxiliary-Center Rule Selection and Consequent Adaptation

Each auxiliary center selects locally supported rules from $K_q^{\mathrm{sh}}$, rather than simply retaining $K^m$, and adapts consequents under shared semantics $\Psi^m$, using main-center priors only for matched antecedents.

**Local Support-Based Rule Selection at the Auxiliary Centers.** For each $\kappa = (S, l_S) \in K_q^{\mathrm{sh}}$, the unnormalized firing strength on auxiliary sample $\mathbf{x}_i^{(a)}$ and local support are

$$h_{i,\kappa}^{(a)} = \prod\nolimits_{j \in S} \mu_{j,l_j}^m (x_{i,j}^{(a)}), \;\; \kappa = (S, l_S), \qquad s^{(a)}(\kappa) = \frac{1}{N_a} \sum\nolimits_{i=1}^{N_a} h_{i,\kappa}^{(a)}. \tag{6}$$

The same order-wise screening uses local supports: $K^{(a)} = \mathsf{HS}(K_q^{\mathrm{sh}}, s^{(a)})$, preserving identity comparability while retaining locally supported rules.

**Lightweight Consequent Transfer at the Auxiliary Centers.** Only the main–auxiliary matched identities $K_{\mathrm{tr}}^{(a)} = K^{(a)} \cap K^m$ receive main-center consequent priors; transfer terms vanish when $K_{\mathrm{tr}}^{(a)} = \emptyset$. For any $\kappa \in K^{(a)}$, define $\tilde{h}_{i,\kappa}^{(a)} = h_{i,\kappa}^{(a)} / \sum_{\kappa' \in K^{(a)}} h_{i,\kappa'}^{(a)}$, $\phi_{i,\kappa}^{(a)} = \tilde{h}_{i,\kappa}^{(a)}$, and $\mathbf{\Phi}^{(a)} = [\phi_{i,\kappa}^{(a)}]_{i=1,\ldots,N_a;\, \kappa \in K^{(a)}} \in \mathbb{R}^{N_a \times |K^{(a)}|}$, with $\mathbf{P}^{(a)} \in \mathbb{R}^{|K^{(a)}| \times C}$. Let $\mathbf{P}_{\mathrm{tr}}^{(a)}$ and $\mathbf{P}_{\mathrm{tr}}^m$ denote the auxiliary- and main-center consequent blocks over $K_{\mathrm{tr}}^{(a)}$. The corresponding average rule-activation vectors are

$$\mathbf{\Gamma}_{\mathrm{tr}}^{(a)} = \frac{1}{N_a} \mathbf{1}_{N_a}^T \mathbf{\Phi}_{\mathrm{tr}}^{(a)}, \quad \mathbf{\Gamma}_{\mathrm{tr}}^m = \frac{1}{N_m} \mathbf{1}_{N_m}^T \mathbf{\Phi}_{\mathrm{tr}}^m, \tag{7}$$

where $\mathbf{\Phi}_{\mathrm{tr}}^{(a)}$ and $\mathbf{\Phi}_{\mathrm{tr}}^m$ are the columns of the auxiliary- and main-center design matrices indexed by $K_{\mathrm{tr}}^{(a)}$, without subsequent renormalization. The rule-transfer consequent-learning objective for the $a$-th auxiliary center is

$$\begin{aligned} \min_{\mathbf{P}^{(a)}} \frac{1}{2}\|\mathbf{Y}^{(a)} - \mathbf{\Phi}^{(a)} \mathbf{P}^{(a)}\|_F^2 + \frac{\lambda_1^{(a)}}{2}\|\mathbf{P}_{\mathrm{tr}}^{(a)} - \mathbf{P}_{\mathrm{tr}}^m\|_F^2 \\ + \frac{\lambda_2^{(a)}}{2}\|\mathbf{P}^{(a)}\|_F^2 + \frac{\lambda_3^{(a)}}{2}\left\|\mathbf{\Gamma}_{\mathrm{tr}}^{(a)} \mathbf{P}_{\mathrm{tr}}^{(a)} - \mathbf{\Gamma}_{\mathrm{tr}}^m \mathbf{P}_{\mathrm{tr}}^m\right\|_F^2, \end{aligned} \tag{8}$$

where $\lambda_1^{(a)}, \lambda_3^{(a)} \geq 0$ and $\lambda_2^{(a)} > 0$. Let $\mathbf{S}^{(a)} \in \{0,1\}^{|K_{\mathrm{tr}}^{(a)}| \times |K^{(a)}|}$ satisfy $\mathbf{P}_{\mathrm{tr}}^{(a)} = \mathbf{S}^{(a)} \mathbf{P}^{(a)}$. Solving the normal equations yields

$$\begin{aligned} \mathbf{P}^{(a)} = \big((\mathbf{\Phi}^{(a)})^T \mathbf{\Phi}^{(a)} + \lambda_2^{(a)} \mathbf{I} + \lambda_1^{(a)} (\mathbf{S}^{(a)})^T \mathbf{S}^{(a)} + \lambda_3^{(a)} (\mathbf{S}^{(a)})^T (\mathbf{\Gamma}_{\mathrm{tr}}^{(a)})^T \mathbf{\Gamma}_{\mathrm{tr}}^{(a)} \mathbf{S}^{(a)}\big)^{-1} \\ \cdot \big((\mathbf{\Phi}^{(a)})^T \mathbf{Y}^{(a)} + \lambda_1^{(a)} (\mathbf{S}^{(a)})^T \mathbf{P}_{\mathrm{tr}}^m + \lambda_3^{(a)} (\mathbf{S}^{(a)})^T (\mathbf{\Gamma}_{\mathrm{tr}}^{(a)})^T \mathbf{\Gamma}_{\mathrm{tr}}^m \mathbf{P}_{\mathrm{tr}}^m\big), \end{aligned} \tag{9}$$

where $\mathbf{I}$ is the identity matrix of size $|K^{(a)}| \times |K^{(a)}|$.

### 3.4 TARGET-CENTER RULE SELECTION, CONSEQUENT FUSION, AND LEARNING

The target center learns the deployment classifier by selecting rules from a source-supported pool using local evidence and integrating matched consequents.

**Source-Aware Target Rule Selection.** Rules with the same $(S, l_S)$ under $\Psi^m$ share an antecedent identity, although their consequents may differ. The source rule repository is defined as $K^{\mathrm{uni}} = K^m \cup \bigcup_{a=1}^{A} K^{(a)}$. Let $\mathcal{U}^{\mathrm{src}} = \{m, 1, \ldots, A\}$, with $K^u$, $\mathbf{P}^u$, $s^u(\kappa)$, $\mathbf{\Gamma}^u = N_u^{-1}\mathbf{1}_{N_u}^\top \mathbf{\Phi}^u$, and $N_u$ denoting, respectively, the rule set, consequent matrix, support function, average activation vector, and sample size of source $u$, where a source index $u = a$ refers to the $a$-th auxiliary center so that $K^u = K^{(a)}$. For any $\kappa \in K^{\mathrm{uni}}$, define the source-retention indicator as $\chi_u(\kappa) = \mathbf{1}\{\kappa \in K^u\}, u \in \mathcal{U}^{\mathrm{src}}$. The source-side support proxy is

$$r^{\mathrm{src}}(\kappa) = \sum\nolimits_{u \in \mathcal{U}^{\mathrm{src}}} \pi_u \chi_u(\kappa) s^u(\kappa), \pi_u \geq 0, \sum\nolimits_{u \in \mathcal{U}^{\mathrm{src}}} \pi_u = 1, \tag{10}$$

where $\pi_u$ is the source weight, defaulting to $\pi_u = N_u / \sum_v N_v$. The score $r^{\mathrm{src}}(\kappa)$ is a zero-padded weighted support over sources and favors rules retained and strongly activated in high-weight centers. Order-wise screening of $K^{\mathrm{uni}}$ by $r^{\mathrm{src}}$, together with the lower-order closure constraint, yields $K_o^{\mathrm{src}}$ for each nonempty order $o$. The resulting pool $K^{\mathrm{src}} = \bigcup_{o=1}^{q} K_o^{\mathrm{src}}$ is used for target-center selection.

Given $K^{\mathrm{src}}$, the target center identifies a locally suitable subfamily without generating new rules. For any $\kappa = (S, l_S) \in K^{\mathrm{src}}$, its unnormalized firing strength on $\mathbf{x}_i^t$ is $h_{i,\kappa}^t = \prod_{j \in S} \mu_{j,l_j}^m(x_{i,j}^t)$, and the local support is $s^t(\kappa) = \frac{1}{N_t} \sum_{i=1}^{N_t} h_{i,\kappa}^t$. Since $N_t$ is usually small, target-side support is combined with the source-side support proxy:

$$\hat{r}^t(\kappa) = \beta_t s^t(\kappa) + (1 - \beta_t) r^{\mathrm{src}}(\kappa), \quad \beta_t = \frac{N_t}{N_t + n_0}, \tag{11}$$

where $n_0 > 0$ is a balancing constant. As $N_t$ increases, the score shifts from source evidence toward target evidence.

Target-side rule identification follows the same source-pool convention: for each nonempty order $o$, the threshold is computed over $K_o^{\mathrm{src}}$ using $\hat{r}^t$, after which the lower-order closure constraint yields $K_o^t$; empty order-wise sets retain no target rules. Thus, $K^t = \bigcup_{o=1}^{q} K_o^t$ is selected jointly by source-side support and target-side activation evidence.

**Partial-Overlap Consequent Fusion and Target Learning.** To accommodate partially overlapping source rule sets, each target rule $\kappa \in K^t$ receives consequent blocks only from sources in $\mathcal{U}_\kappa = \{u \in \mathcal{U}^{\mathrm{src}} : \kappa \in K^u\}$. To weight these contributions by rule-usage consistency, define the zero-padded source support profile and the target relevance profile over $K^t$ as $\bar{\mathbf{s}}_{K^t}^u = [\chi_u(\kappa) s^u(\kappa)]_{\kappa \in K^t} \in \mathbb{R}^{|K^t|}, \hat{\mathbf{r}}_{K^t}^t = [\hat{r}^t(\kappa)]_{\kappa \in K^t} \in \mathbb{R}^{|K^t|}$. The antecedent-support-profile discrepancy is $d_u = \|\bar{\mathbf{s}}_{K^t}^u - \hat{\mathbf{r}}_{K^t}^t\|_2^2$. For each $\kappa \in K^t$ and $u \in \mathcal{U}_\kappa$, the availability-conditioned fusion weight is

$$\omega_{u,\kappa} = \frac{\exp(-d_u/\tau_\omega)}{\sum_{v \in \mathcal{U}_\kappa} \exp(-d_v/\tau_\omega)}, \ \sum\nolimits_{u \in \mathcal{U}_\kappa} \omega_{u,\kappa} = 1, \tag{12}$$

where $\tau_\omega > 0$ is the temperature parameter. Let $\mathbf{p}_\kappa^u \in \mathbb{R}^C$ denote the consequent block of source $u$ associated with rule $\kappa$. The rule-level fusion prior is $\mathbf{p}_{\mathrm{fus},\kappa}^t = \sum_{u \in \mathcal{U}_\kappa} \omega_{u,\kappa} \mathbf{p}_\kappa^u$. Stacking $\{\mathbf{p}_{\mathrm{fus},\kappa}^t\}_{\kappa \in K^t}$ in the canonical rule order gives the fusion-prior matrix $\mathbf{P}_{\mathrm{fus}}^t \in \mathbb{R}^{|K^t| \times C}$. The rule-level fusion prior regularizes individual rule blocks, whereas the target center may also require stabilization at the response level. Reusing $d_u$, the source-level weight is

$$\alpha_u = \frac{\exp(-d_u/\tau_\omega)}{\sum_{v \in \mathcal{U}^{\mathrm{src}}} \exp(-d_v/\tau_\omega)}, \ u \in \mathcal{U}^{\mathrm{src}}, \sum\nolimits_{u \in \mathcal{U}^{\mathrm{src}}} \alpha_u = 1. \tag{13}$$

Compared with $\omega_{u,\kappa}$ in (12), $\alpha_u$ is a heuristic source weight based on support-profile compatibility with the target-selected rule structure. Let $\widetilde{\mathbf{\Gamma}}_{K^t}^u$ and $\widetilde{\mathbf{P}}_{K^t}^u$ denote the $K^t$-aligned, zero-padded source activation mean and consequent matrix. The response-level anchor is

$$\mathbf{r}_{\mathrm{fus}}^t = \sum\nolimits_{u \in \mathcal{U}^{\mathrm{src}}} \alpha_u \widetilde{\mathbf{\Gamma}}_{K^t}^u \widetilde{\mathbf{P}}_{K^t}^u. \tag{14}$$

Here $\mathbf{r}_{\mathrm{fus}}^t \in \mathbb{R}^{1\times C}$ complements $\mathbf{P}_{\mathrm{fus}}^t$. For any $\kappa \in K^t$, the target-center normalized firing strength and rule-level feature are $\tilde{h}_{i,\kappa}^t = h_{i,\kappa}^t / \sum_{\kappa' \in K^t} h_{i,\kappa'}^t$, and $\phi_{i,\kappa}^t = \tilde{h}_{i,\kappa}^t$, respectively. Stacking $\phi_{i,\kappa}^t$ gives $\mathbf{\Phi}^t = [\phi_{i,\kappa}^t]_{i=1,\ldots,N_t;\, \kappa\in K^t} \in \mathbb{R}^{N_t\times|K^t|}$, with consequent matrix $\mathbf{P}^t \in \mathbb{R}^{|K^t|\times C}$. The target-center average rule-activation vector is $\mathbf{\Gamma}^t = \frac{1}{N_t}\mathbf{1}_{N_t}^\top \mathbf{\Phi}^t \in \mathbb{R}^{1\times|K^t|}$. The final target-center consequent is learned by solving the following convex quadratic problem:

$$\min_{\mathbf{P}^t} \frac{1}{2}\|\mathbf{Y}^t - \mathbf{\Phi}^t\mathbf{P}^t\|_F^2 + \frac{\lambda_1^t}{2}\|\mathbf{P}^t - \mathbf{P}_{\mathrm{fus}}^t\|_F^2 + \frac{\lambda_2^t}{2}\|\mathbf{P}^t\|_F^2 + \frac{\lambda_3^t}{2}\|\mathbf{\Gamma}^t\mathbf{P}^t - \mathbf{r}_{\mathrm{fus}}^t\|_F^2, \tag{15}$$

where $\lambda_1^t$, $\lambda_2^t$, and $\lambda_3^t$ are nonnegative weights. When $\lambda_1^t + \lambda_2^t > 0$, problem (15) has the closed-form solution

$$\mathbf{P}^t = \left((\mathbf{\Phi}^t)^\top\mathbf{\Phi}^t + (\lambda_1^t + \lambda_2^t)\mathbf{I} + \lambda_3^t(\mathbf{\Gamma}^t)^\top\mathbf{\Gamma}^t\right)^{-1} \cdot \left((\mathbf{\Phi}^t)^\top\mathbf{Y}^t + \lambda_1^t\mathbf{P}_{\mathrm{fus}}^t + \lambda_3^t(\mathbf{\Gamma}^t)^\top\mathbf{r}_{\mathrm{fus}}^t\right), \tag{16}$$

where $\mathbf{I}$ is the identity matrix of size $|K^t| \times |K^t|$. The supplementary method details and Table S2 describe key-based alignment, zero padding, and the complete training sequence.

# 4 Experimental Results

## 4.1 Datasets and Experimental Setup

We evaluate MCRT-TSK on 18 UCI tabular classification datasets (Kelly et al., 2026) (10 binary and 8 multi-class; Table S3 in the supplementary material). Experiments were conducted in MATLAB R2024a on a Windows workstation with an Intel(R) Core(TM) Ultra 9 285K CPU and 64 GB RAM, without GPU acceleration or external frameworks. Results are averaged over ten runs with seeds 2–11. The main comparison reports target-center accuracy (ACC) against TL-SSL-TSK (Jiang et al., 2017), ETTL-TSK-FS (Deng et al., 2018), FL-TSK-FRBSs (Corcuera Bárcena et al., 2022), MKTC-R0T (Wang et al., 2024), MIP-TRL-FS (Li et al., 2024a), KCF-LSG-MTSK (Wang et al., 2026), and MTSK-DA (Yao et al., 2026). Supplementary Tables S5–S6 additionally report target-center Macro-F1 and G-mean (GM) under the ACC-selected configurations. Target-only 0-TSK and 1-TSK are diagnostic reference points rather than transfer baselines. In each of the ten runs, datasets are stratified into main (30%), two auxiliary (25% each), and target (20%) centers, each further split 80%/20% into training/test sets. With target-training sizes ranging from 24 to 877, our benchmark focuses on limited target supervision and structural mismatch rather than few-shot learning or natural institutional shifts (finite-sample audits in Figs. S4–S5). We report target-test means and standard deviations, utilizing five-fold cross-validation where applicable. Default parameters are $q = 2$, $L_{\max} = 3$, $\boldsymbol{\rho} = (0.8, 0.9)$, $n_0 = 20$, and $\tau_\omega = 0.1$; tuning grids include $\tau_{\min} \in \{10^{-4}, 10^{-3}\}$, $\eta \in \{0.5, 1, 2\}$, and regularizers in $\{10, 1, 0.01\}$. The supplementary material gives the complete settings.

## 4.2 Performance Comparison

Table 1 reports target-test ACC for MCRT-TSK and nine comparators. The multi-class panel includes only methods whose released implementations support the evaluated multi-class tasks. TO denotes a run exceeding 120 hours; unsupported and timeout entries are not imputed.

MCRT-TSK is best or tied-best on 16 of 18 datasets, covering all eight multi-class tasks and eight binary tasks. Among methods with valid outputs on every multi-class dataset, it achieves the strongest ACC results; Ban and Spa favor 1-TSK and FL-TSK-FRBSs, respectively. These results provide evidence of competitive performance under the controlled simulated-center protocol.

## 4.3 Statistical Significance Analysis

After excluding status-only entries, the Friedman test on the strict eight-dataset complete subset gives MCRT-TSK the best average rank (2.06) and rejects the equal-performance null hypothesis ($p = 4.50 \times 10^{-5}$). Wilcoxon tests on each valid pairwise intersection (9–18 datasets) yield positive mean ACC margins against all nine comparators, with seven comparisons remaining significant after Holm correction (Fig. 2(A–B); Supplementary Tables S7–S9).

| Binary classification | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | Son | Hea | Bre | Aus | Blo | Dia | Qsa | DiR | Ban | Spa |
| 0-TSK | 53.75 (18.68) | 78.33 (9.78) | 93.18 (2.40) | 82.96 (8.41) | 75.67 (7.04) | 68.39 (7.87) | 79.52 (7.20) | 61.30 (5.59) | 98.70 (1.76) | 84.70 (3.60) |
| 1-TSK | 50.00 (18.63) | 55.83 (15.24) | 52.73 (20.91) | 73.70 (10.25) | 34.67 (18.34) | 60.97 (10.36) | 42.86 (12.90) | 52.61 (9.27) | **99.44 (1.25)** | 73.84 (18.01) |
| TL-SSL-TSK | TO | 77.50 (11.82) | 95.45 (4.29) | 86.67 (5.00) | 78.67 (6.32) | 70.97 (5.89) | **82.86 (7.68)** | 63.70 (5.98) | 96.85 (1.76) | 88.22 (2.51) |
| ETTL-TSK-FS | 76.25 (20.79) | 79.17 (9.82) | 94.55 (4.69) | 88.52 (6.16) | 75.33 (10.80) | 70.97 (6.80) | 82.38 (5.17) | 63.26 (6.52) | 96.67 (2.59) | 83.68 (2.45) |
| FL-TSK-FRBSs | 53.75 (13.24) | 59.17 (16.41) | 73.64 (7.96) | 55.56 (0.00) | 79.67 (5.08) | 72.58 (6.13) | 78.10 (5.24) | 62.17 (8.27) | 96.85 (2.32) | **90.38 (1.50)** |
| MKTC-R0T | 63.75 (14.97) | 74.17 (12.08) | 93.64 (3.18) | 82.59 (6.54) | 75.00 (6.14) | 70.00 (7.61) | 74.76 (5.52) | 60.43 (8.50) | 98.70 (1.25) | 88.05 (3.10) |
| MIP-TRL-FS | 63.75 (16.08) | 74.17 (14.93) | 78.64 (13.39) | 72.22 (10.07) | 65.67 (13.79) | 60.32 (6.46) | 71.67 (9.75) | 54.57 (5.93) | 73.15 (14.59) | 81.51 (3.10) |
| MTSK-DA | 65.00 (17.48) | 80.00 (10.54) | 95.45 (2.14) | 89.26 (6.64) | 69.67 (7.28) | 66.45 (5.93) | 80.95 (6.92) | 61.74 (6.58) | 98.52 (1.91) | TO |
| KCF-LSG-MTSK | 80.00 (12.08) | 80.83 (10.43) | 95.00 (4.52) | 85.19 (8.00) | 72.67 (11.53) | 72.90 (8.76) | 80.00 (7.95) | 63.91 (5.63) | 98.33 (2.22) | TO |
| MCRT-TSK (ours) | **85.00 (11.49)** | **88.33 (7.03)** | **96.82 (3.07)** | **90.74 (3.60)** | **82.33 (4.17)** | **74.52 (4.15)** | **82.86 (2.19)** | **65.22 (2.29)** | 95.93 (3.24) | 87.14 (1.08) |

| Multi-class classification | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method | Iri | Win | Con | Car | WiQ | WiD | Wav | Pag |
| 0-TSK | 91.67 (11.79) | 92.86 (12.14) | 46.90 (7.66) | 74.49 (2.83) | 56.46 (2.31) | 51.08 (2.74) | 81.50 (1.76) | 91.50 (0.53) |
| 1-TSK | 71.67 (20.86) | 52.86 (21.35) | 50.52 (7.13) | 77.39 (1.56) | 56.57 (3.62) | 52.62 (4.39) | 80.45 (2.10) | 93.68 (0.58) |
| MIP-TRL-FS | 68.33 (35.53) | 52.86 (30.15) | 37.76 (3.19) | 53.91 (10.17) | 51.77 (4.37) | 45.69 (5.56) | 74.80 (3.31) | 91.00 (1.68) |
| MTSK-DA | 91.67 (11.79) | 95.71 (9.64) | 52.76 (4.54) | 60.87 (21.10) | TO | TO | TO | TO |
| KCF-LSG-MTSK | 91.67 (11.79) | 97.14 (6.02) | 50.52 (6.61) | 55.07 (15.25) | 57.57 (4.55) | 49.74 (3.49) | 82.90 (3.19) | TO |
| MCRT-TSK (ours) | **100.00 (0.00)** | **100.00 (0.00)** | **53.28 (1.71)** | **78.84 (3.36)** | **59.50 (2.08)** | **55.08 (2.14)** | **84.55 (1.01)** | **94.00 (0.47)** |

Table 1: Performance comparison of MCRT-TSK and nine comparators on 18 datasets (ACC↑, %; mean above standard deviation in parentheses). Boldface and underlining indicate the best and second-best valid results, respectively. The multi-class panel omits TL-SSL-TSK, ETTL-TSK-FS, FL-TSK-FRBSs, and MKTC-R0T, whose released implementations do not support these multi-class tasks. TO: timeout ($>$ 120 hours per run).

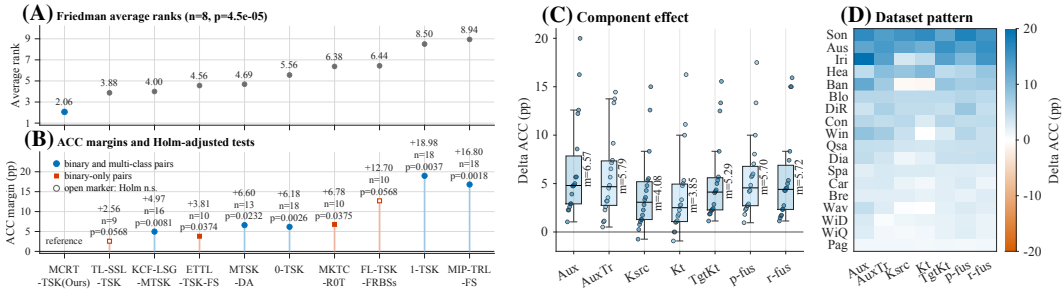


Figure 2: ACC statistical summary ($\alpha$ = 0.05) and component-removal evidence under $q = 2$. **(A)** Friedman ranks on the strict complete all-method subset; **(B)** paired MCRT-TSK margins and Holm-adjusted Wilcoxon $p$-values; **(C)** component distributions; and **(D)** dataset-level deltas. Full-minus-variant ACC deltas are in percentage points; positive values indicate degradation.

### 4.4 Ablation Study

Under $q = 2$, seven same-protocol ablations remove, respectively, the auxiliary centers, auxiliary transfer, source-pool screening $K^{\mathrm{src}}$, target filtering for $K^t$, mixed source–target scoring for $K^t$, $P^t_{\mathrm{fus}}$, or $r^t_{\mathrm{fus}}$; the exact switches are listed in Supplementary Table S10. Figure 2(C) shows the largest mean losses after removing auxiliary data/transfer (6.57/5.79 pp) or the two fusion priors

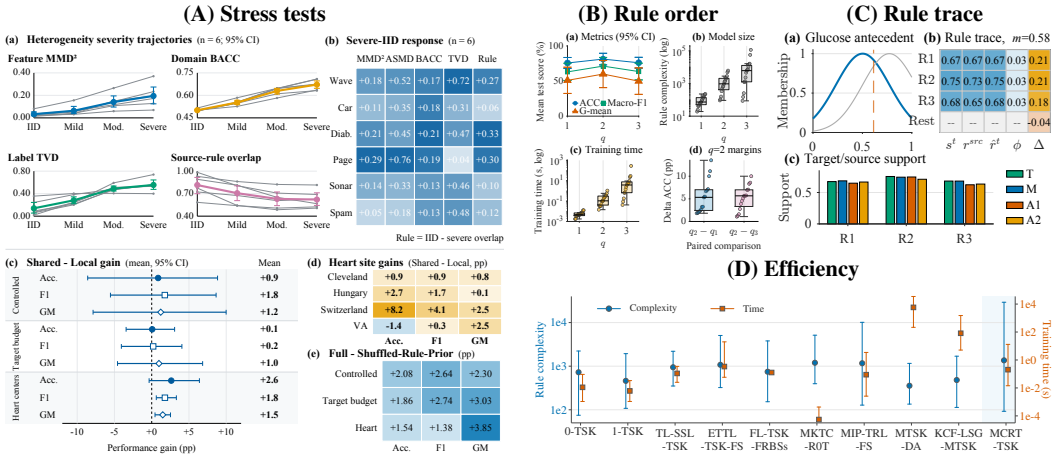


Figure 3: Stress tests and model diagnostics. **(A)** Stress tests: (a) heterogeneity trajectories (95% CIs); (b) severe-minus-IID diagnostics (Rule: IID-minus-severe overlap); (c) Shared-Antecedent minus Local-Target mean gains (95% CIs); (d) Heart-site gains; and (e) Full minus Shuffled-Rule-Prior mean gains (pp). **(B)** Maximum rule order $q$: (a) predictive metrics (95% CIs), (b) rule complexity, (c) training time, and (d) paired ACC margins. **(C)** Diabetes rule trace: (a) a representative antecedent, (b) support, mixed relevance, firing, and margin contribution (margin = 0.58), and (c) matched target/source supports ($T$, $M$, $A_1$, $A_2$). **(D)** Rule complexity and training time across valid datasets: markers show medians and segments show min–max ranges on log scales.

(5.70/5.72 pp), with dataset-level deltas shown in panel (D). Target-only selection and screening removals lose 5.29, 4.08, and 3.85 pp on average. Near-zero or reversed effects on some datasets indicate aggregate, not universal, component contributions.

### 4.5 Heterogeneity, Partial Rule Overlap, and Mechanism Stress Tests

We first isolate antecedent transfer under controlled feature and label heterogeneity, three target-label budgets, and four natural UCI Heart sites. Distributional diagnostics rise and source-rule overlap falls with severity (panels (a,b) of Fig. 3(A)); the Heart rotations likewise show domain BACC of 0.699–0.845 and partial source-rule overlap (Jaccard 0.632–0.659). Shared-Antecedent, which excludes source-consequent priors, yields positive mean ACC, Macro-F1, and G-mean gains over Local-Target in all three families; the family-level and Heart-site gains are summarized in panels (c–d) of Fig. 3(A).

We then break only the rule-identity–consequent correspondence: Shuffled-Rule-Prior preserves the target rules, data, prior size and norm, response anchor, and regularization, but permutes rule-prior rows within each order. As summarized in panel (e) of Fig. 3(A), Full yields positive mean margins over this control in all nine task-family–metric summaries, ranging from $+1.38$ to $+3.85$ pp. Because this control changes only the correspondence between rule identity and consequent prior, the consistent positive margins provide mechanism-specific evidence for identity-preserving consequent reuse. Full block-level results appear in the supplementary material.

## 5 Analysis and Discussion

### 5.1 Time Complexity Analysis

Let $U = A + 1$, $L_\Sigma = \sum_j L_j$, $\bar{L} = \max_j L_j \le L_{\max}$, and $R_\mathcal{K} = \sum_{\kappa \in \mathcal{K}} \mathrm{ord}(\kappa)$. Lloyd partitioning costs $O(dI_k N_m L_{\max}^2)$, while the shared candidate count is $M_q = \sum_{r=1}^{q} \sum_{|S|=r} \prod_{j \in S} L_j \le \sum_{r=1}^{q} \binom{d}{r} \bar{L}^r$ and identity construction costs $O(qM_q \log M_q)$. For $N$ samples and retained set $\mathcal{K}$, firing costs $O(NL_\Sigma + NR_\mathcal{K})$; fitting $M = |\mathcal{K}|$ consequents costs $O(NM^2 + NMC + M^3 + M^2C)$ in addition to firing. Source merging/alignment adds $O(M_\Sigma \log M_\Sigma + \sum_u N_u(M_{\mathrm{uni}} + M_t) +$

$U(M_{\text{uni}} + M_t)C)$, and the target fit has the same retained-rule form with $M_t$. Prediction costs $O(N_e L_\Sigma + N_e R_{K^t} + N_e M_t C)$. Thus, training is governed by $M_q$, retained-rule cardinalities, alignment, and the largest linear solve; $\mathrm{RC}(K) = \sum_{\kappa \in K} C(\mathrm{ord}(\kappa) + 1)$ is a structural-size diagnostic, not a runtime measure.

### 5.2 Sensitivity to the Maximum Rule Order

As shown in Fig. 3(B), across the 15 datasets with valid results for all tested $q$ settings (excluding Qsa, Son, and Spa, for which $q = 3$ produced no valid results), $q = 2$ yielded the highest mean ACC (81.28%), Macro-F1 (71.29%), and G-mean (59.95%), and outperformed $q = 1$ and $q = 3$ in ACC on every dataset. Relative to $q = 2$, $q = 3$ increased the retained-rule complexity diagnostic by $10.1\times$ and training time by $105.6\times$, without improving ACC. Because candidate growth is already pronounced at $q = 3$, values above 3 were outside the tested range. These observations support $q = 2$ as the default among the evaluated settings.

### 5.3 Empirical Efficiency Analysis

Figure 3(D) compares implementation-aligned retained size and training time, not a universal complexity ranking. MCRT-TSK has median and maximum rule complexities of 1361.4/29223.4 and median and maximum training times of 0.2022/13.061 s over 18 datasets. It is slower than target-only diagnostics but markedly faster than heavier transfer baselines MTSK-DA (median 5793.40 s over 13 valid datasets) and KCF-LSG-MTSK (median 81.510 s over 16 valid datasets), indicating moderate cost despite second-order matched-rule screening. Complete matrices and valid-result coverage appear in the supplementary material.

### 5.4 Rule-Level Interpretability Analysis

MCRT-TSK preserves transferred knowledge as antecedent-matched rule blocks and permits linguistic rendering when feature semantics are available. Figure 3(C) traces a Diabetes prediction through $K^{\text{uni}} \to K^{\text{src}} \to K^t$, reporting target support, the source-side support proxy, mixed relevance, sample firing, and margin contribution. With $N_t = 123$ and $n_0 = 20$, $\beta_t = 0.8601$, so target evidence dominates screening while source evidence remains in the priors. The selected sample is correctly classified as class 2 with raw scores $(0.2105, 0.7864)$. Its leading rule—"very low" pregnancies and "medium" glucose—has $s^t = 0.674$, $r^{\text{src}} = 0.668$, $\hat{r}^t = 0.673$, $\phi_\kappa(\mathbf{x}_*) = 0.030$, and margin contribution $0.2148$; the first three positive rules explain most of the $0.5759$ margin. The complete antecedent, source-fusion, response-anchor, and decision decomposition is detailed in the supplementary material.

The trace separates shared semantics, local evidence, and transferred consequents. Each antecedent identity is defined by a feature subset and fuzzy-set index assignment under the shared semantics, preserving correspondence across different local rule subsets. Normalized sample firing strengths combine with learned target consequents to determine prediction contributions, whereas matched source consequents form the rule-level prior. The displayed target/source supports connect the selected prediction to this aligned rule structure.

## 6 Conclusion

MCRT-TSK aligns partially overlapping local rule bases through shared antecedent identities, balances source support with target activation during rule selection, and fuses only matched consequent blocks in a closed-form target learner. The primary benchmark supports the complete model under the simulated-center protocol. Stress tests verify distributional and rule-structural heterogeneity, show positive mean gains for shared antecedent transfer, and favor identity-preserving over shuffled consequent correspondence in all nine task-family–metric summaries. These results identify antecedent identity as both a transferable structural interface and a necessary key for consequent reuse. Future work will consider acquisition-induced shifts, severe class imbalance, mismatch-aware prior calibration, and non-identical feature spaces.

### AI USE STATEMENT

Generative AI tools were used only for language editing to improve the grammar, phrasing, and readability of the manuscript. All AI-assisted edits were reviewed by the authors. No generative AI tools were used for the research methodology, experiments, analysis, or interpretation of results. The authors take full responsibility for the final content of the paper.

## REPRODUCIBILITY STATEMENT

The method specification and experimental setup appear in the main text. The appendices provide implementation conventions, hyperparameter settings, complete results, and additional analyses. The accompanying supplementary package contains the MATLAB code, data, and reproduction instructions used for the reported experiments.

## REFERENCES

José Luis Corcuera Bárcena, Pietro Ducange, Alessio Ercolani, Francesco Marcelloni, and Alessandro Renda. An approach to federated learning of explainable fuzzy regression models. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8. IEEE, 2022. doi: 10.1109/FUZZ-IEEE55066.2022.9882881.

Zhaohong Deng, Peng Xu, Lixiao Xie, Kup-Sze Choi, and Shitong Wang. Transductive joint-knowledge-transfer tsk fs for recognition of epileptic eeg signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(8):1481–1494, 2018. doi: 10.1109/TNSRE.2018.2850308.

Wei Guo, Fuzhen Zhuang, Xiao Zhang, Yiqi Tong, and Jin Dong. A comprehensive survey of federated transfer learning: Challenges, methods and applications. *Frontiers of Computer Science*, 18(6):186345, 2024. doi: 10.1007/s11704-024-40065-x.

J.-S. R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993. doi: 10.1109/21.256541.

Yizhang Jiang, Dongrui Wu, Zhaohong Deng, Pengjiang Qian, Jun Wang, Guanjin Wang, Fu-Lai Chung, Kup-Sze Choi, and Shitong Wang. Seizure classification from eeg signals using transfer learning, semi-supervised learning and tsk fuzzy system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2270–2284, 2017. doi: 10.1109/TNSRE.2017.2748388.

Markelle Kelly, Rachel Longjohn, and Kolby Nottingham. The uci machine learning repository. Available at `https://archive.ics.uci.edu`, 2026. Accessed: 2026-05-20.

Andong Li, Zhaohong Deng, Wei Zhang, Zhiyong Xiao, Kup-Sze Choi, Yueying Liu, Shudong Hu, and Shitong Wang. Multiview transfer representation learning with tsk fuzzy system for eeg epilepsy detection. *IEEE Transactions on Fuzzy Systems*, 32(1):38–52, 2024a. doi: 10.1109/TFUZZ.2023.3288479.

Jingjing Li, Zhekai Yu, Zhaojia Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5743–5762, 2024b. doi: 10.1109/TPAMI.2024.3370978.

Yuchen Li, Fu-Lai Chung, Yusuke Nojima, and Shitong Wang. Fully interpretable gaussian centralized tsk fuzzy classifier from probabilistic perspective: Concepts, output-stability-based learning, and ensemble. *IEEE Transactions on Fuzzy Systems*, 33(11):3967–3980, 2025. doi: 10.1109/TFUZZ.2025.3600911.

Guangzhi Ma, Jie Lu, and Guangquan Zhang. Multisource domain adaptation with interval-valued target data via fuzzy neural networks. *IEEE Transactions on Fuzzy Systems*, 32(5):3094–3106, 2024. doi: 10.1109/TFUZZ.2024.3367456.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Aarti Singh and Jerry Zhu (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi: 10.1109/TKDE.2009.191.

Michio Sugeno and Gi Taek Kang. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28(1):15–33, 1988. doi: 10.1016/0165-0114(88)90113-3.

Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(1):116–132, 1985. doi: 10.1109/TSMC.1985.6313399.

Hongpeng Tian, Zuowei Zhang, and Weiping Ding. Incomplete data transfer calibration classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(5):3244–3256, 2025. doi: 10.1109/TETCI.2024.3449902.

Chuang Wang, Pengjiang Qian, Zhihuang Wang, Weiwei Cai, Jian Yao, Yi-Zhang Jiang, X. Yan, and Wenjun Hu. Multicenter knowledge transfer calibration with rapid zeroth-order tsk fuzzy system for small sample epileptic eeg signals. *IEEE Transactions on Fuzzy Systems*, 32(11):6224–6236, 2024. doi: 10.1109/TFUZZ.2024.3443878.

Chuang Wang, Pengjiang Qian, Weiwei Cai, Jian Yao, Yi-Zhang Jiang, Wenjun Hu, Eddie-Yin-Kwee Ng, and Shitong Wang. Knowledge calibration fusion and label space graph regularization-based multicenter fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 34(4):1349–1360, 2026. doi: 10.1109/TFUZZ.2026.3658901.

L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992. doi: 10.1109/21.199466.

Peng Xu, Zhaohong Deng, Jun Wang, Qun Zhang, Kup-Sze Choi, and Shitong Wang. Transfer representation learning with tsk fuzzy system. *IEEE Transactions on Fuzzy Systems*, 29(3):649–663, 2021. doi: 10.1109/TFUZZ.2019.2958299.

Jian Yao, Weiwei Fu, Pengjiang Qian, Chuang Wang, Xin Guo, Jun Liu, and Zhanjun Zhang. Mtskda: Interpretable multicenter transfer via discriminative alignment and attention-guided graph regularization. *IEEE Transactions on Fuzzy Systems*, 34(7):2211–2224, 2026. doi: 10.1109/TFUZZ.2026.3685305.

Yuanpeng Zhang, Guanjin Wang, Ta Zhou, Ge Ren, Saikit Lam, Weiping Ding, and Jing Cai. Deep reconciled and self-paced tsk fuzzy system ensemble for imbalanced data classification: Architecture, interpretability, and theory. *IEEE Transactions on Fuzzy Systems*, 32(11):6185–6198, 2024. doi: 10.1109/TFUZZ.2024.3442821.

Sicheng Zhao, Hui Chen, Hu Huang, Pengfei Xu, and Guiguang Ding. More is better: Deep domain adaptation with multiple sources. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 8354–8362, 2024. doi: 10.24963/ijcai.2024/923.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1): 43–76, 2021. doi: 10.1109/JPROC.2020.3004555.

# A Supplementary Related Work and Background

## A.1 TSK Fuzzy Systems

Takagi–Sugeno–Kang (TSK) fuzzy systems represent nonlinear mappings through a collection of fuzzy IF–THEN rules with local consequent models (Takagi & Sugeno, 1985; Sugeno & Kang, 1988). For an input $\mathbf{x} = (x_1, \ldots, x_d)^T$, the $k$-th first-order TSK rule is commonly written as

$$\text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_d \text{ is } A_d^k \qquad \text{Then } f^k(\mathbf{x}) = p_0^k + \sum\nolimits_{i=1}^{d} p_i^k x_i,$$

where $A_j^k$ is the fuzzy set assigned to the $j$-th feature and $f^k(\mathbf{x})$ is the local consequent function. In a zero-order TSK model, the consequent reduces to a constant or, for classification, a rule-specific class-response vector. The standard TSK model therefore comprises fuzzy antecedent partitions, rule firing strengths, and local consequent responses.

Given a retained rule set $K$, the firing and normalized firing strengths are computed as

$$h_k(\mathbf{x}) = \prod\nolimits_{j=1}^{d} \mu_{A_j^k}(x_j), \qquad \tilde{h}_k(\mathbf{x}) = \frac{h_k(\mathbf{x})}{\sum_{\ell \in K} h_\ell(\mathbf{x})}.$$

The final output is obtained by normalized rule aggregation,

$$f(\mathbf{x}) = \sum\nolimits_{k \in K} \tilde{h}_k(\mathbf{x}) f^k(\mathbf{x}).$$

Early TSK studies established fuzzy identification and structure learning, while later rule-learning and neuro-fuzzy models showed how antecedents and consequents can be estimated from data (Wang & Mendel, 1992; Jang, 1993). These properties explain why TSK systems are widely used when nonlinear prediction and interpretable rule structure are both desired.

When the standard single-center mechanism is extended to multiple centers, the antecedent space can grow rapidly. If the $j$-th feature has $L_j$ fuzzy sets, a full-order rule base contains $\prod_{j=1}^{d} L_j$ antecedents; under maximum rule order $q$, the candidate count is

$$\sum\nolimits_{r=1}^{q} \sum\nolimits_{|S|=r} \prod\nolimits_{j \in S} L_j.$$

Practical multi-center TSK learning must therefore determine which antecedent identities are locally supported before reusing their consequents. MCRT-TSK addresses this issue through cross-center identity matching, local support screening, and source-restricted consequent fusion without changing the underlying TSK rule model.

## A.2 Transfer, Domain Adaptation, and Federated Learning

Transfer learning provides the general setting in which a target learner benefits from knowledge acquired in related domains or tasks (Pan & Yang, 2010; Zhuang et al., 2021). Domain adaptation instantiates this idea under distribution shift and has recently moved beyond the classical single-source assumption. Source-free domain adaptation studies the privacy-aware case in which source data are unavailable during target adaptation (Li et al., 2024b), while the multi-source domain adaptation literature explicitly models heterogeneous sources and examines source weighting, representation alignment, intermediate-domain construction, and prediction fusion under non-identical source distributions (Zhao et al., 2024). These developments are closely related in motivation to multi-center learning because they address scarce target supervision, source heterogeneity, and negative transfer.

Federated learning tackles a complementary constraint by training collaboratively without sharing raw data (McMahan et al., 2017). Federated transfer learning combines federated optimization with transfer learning to address heterogeneous feature spaces, non-IID clients, system heterogeneity, incremental data, and label scarcity (Guo et al., 2024). The exchanged or adapted objects are typically feature representations, model parameters, gradients, logits, prototypes, or pseudo-labels; these objects do not by themselves specify correspondence between antecedent-defined fuzzy regions across centers.

The proposed multi-center TSK setting differs from standard domain adaptation and federated transfer learning in two respects. First, the transfer unit is a TSK rule block comprising an antecedent identity, firing statistics, and consequent parameters rather than a whole model or latent representa-

tion. Second, sources may retain different rule subsets, so target learning restricts consequent fusion to semantically matched and locally supported antecedents.

### A.3 Multi-Center TSK Fuzzy Transfer

Recent fuzzy and TSK-based methods have improved cross-center and multi-source learning. TSK transfer representation learning (Xu et al., 2021) supports source adaptation while retaining interpretable local structures. MKTC-style multi-center calibration (Wang et al., 2024) uses zero-order TSK models for small-sample target deployment and knowledge integration. KCF-LSG-MTSK (Wang et al., 2026) combines calibration fusion, projected distribution alignment, and label-space graph regularization. MTSK-DA (Yao et al., 2026) uses discriminative statistic alignment and attention-guided graph regularization.

Table S1: Positioning of MCRT-TSK against representative TSK/fuzzy transfer methods.

| Method | Transfer signal | Main strength | Primary distinction from MCRT-TSK |
|---|---|---|---|
| TSK transfer representation learning | Source-adapted TSK representation and consequents | Connects TSK interpretability with transfer learning | Focuses on representation transfer rather than multi-center partial rule overlap and rule-identity-matched consequent fusion |
| MKTC-R0T | Center-level knowledge calibration with rapid zero-order TSK learning | Efficient small-sample deployment and multi-center knowledge integration | Uses center-level calibration rather than a shared bounded-order identity space with source–target support screening |
| KCF-LSG-MTSK | Calibration fusion, projected distribution alignment, and label-space graph regularization | Uses complementary distributional and label-space structure across centers | Uses calibration, distributional, and graph signals rather than retained rule identities with partial overlap |
| MTSK-DA | Discriminative statistic alignment and attention-guided graph regularization | Promotes intra-class compactness and inter-class separation across centers | Uses distributional/graph alignment rather than matched identities and source-restricted consequent fusion; its optimization structure also differs |
| MCRT-TSK | Shared rule identities, local fuzzy support, and matched consequent fusion | Transfers semantically matched and locally supported rule blocks | Uses shared identities, local support screening, and matched consequent fusion for partially overlapping retained rule sets |

MTSK-DA is the most directly related comparator because it is a discriminative multi-center TSK method rather than a generic domain-adaptation method. It combines discriminative statistic alignment with attention-guided graph regularization, whereas MCRT-TSK uses shared antecedent identities and support-based rule screening.

The two methods differ primarily in their transfer interfaces and optimization structures: MTSK-DA imposes distributional and graph constraints through coupled optimization, whereas MCRT-TSK performs rule-identity-matched consequent fusion followed by closed-form target estimation.

MCRT-TSK complements these methods by using shared TSK antecedent identities as the transfer interface. A main center defines the antecedent coordinate system, each center screens locally supported rules, and consequent fusion is restricted to sources retaining the same target rule.

# B Supplementary Method Details

## B.1 Feature Partitioning and Hierarchical Screening

For the feature-level partition criterion in the main text, $\mathcal{C}_{j,L}$ is returned by the deterministic one-dimensional Lloyd routine applied to the sorted main-center feature values, and $D_j(L)$ is the resulting within-cluster distortion. Accordingly, $D_j(L)$ denotes the solution attained by that routine; neither the implementation nor the analysis assumes a globally optimal one-dimensional quantizer. The same routine is used across repetitions, and the selected distortion and ordered center set are retained for subsequent training.

The following definitions specify the screening operator used in the main text. Let $\mathcal{K}$ be a candidate rule set whose rules have orders $|S| \le q$, and let $\zeta(\kappa) \ge 0$ be a rule score. The screening parameters satisfy $0 < \rho_1 \le \cdots \le \rho_q \le 1$ and $\tau_{\min} > 0$. For a nonempty multiset $\mathcal{Z}$ of $M$ scores sorted as $z_{(1)} \le \cdots \le z_{(M)}$, let $\alpha_o = 1 - \rho_o$, $p_o = 1 + (M-1)\alpha_o$, $\ell_o = \lfloor p_o \rfloor$, and $u_o = \lceil p_o \rceil$. The empirical quantile $Q_{1-\rho_o}(\mathcal{Z})$ is

$$Q_{1-\rho_o}(\mathcal{Z}) = \begin{cases} z_{(\ell_o)}, & \ell_o = u_o, \\ (u_o - p_o) z_{(\ell_o)} + (p_o - \ell_o) z_{(u_o)} & \ell_o < u_o. \end{cases}$$

This interpolation rule matches the implementation used in the experiments. Rules whose scores equal the threshold are retained, so ties may retain slightly more than the nominal fraction. For a rule $\kappa = (S, l_S)$ with $|S| = o$, define its inherited $(o-1)$-order subrule set as

$$\mathrm{sub}_{o-1}(\kappa) = \{(T, l_T) : T \subset S,\ |T| = o - 1,\ l_T = l_S|_T\}.$$

For main- and auxiliary-center support-based rule screening, the empirical quantile threshold is computed over the recursively eligible candidate set at that order. For order $o = 1$, define

$$C_1 = \{\kappa \in \mathcal{K} : |S| = 1\}, \qquad \mathcal{R}_1 = \{\zeta(\kappa) : \kappa \in C_1\}.$$

If $C_1$ is empty, set $K_1 = \emptyset$. Otherwise, let

$$\tau_1 = \max\{Q_{1-\rho_1}(\mathcal{R}_1), \tau_{\min}\}, \qquad K_1 = \{\kappa \in C_1 : \zeta(\kappa) \ge \tau_1\}.$$

For each higher order $o = 2, \ldots, q$, define

$$C_o = \{\kappa \in \mathcal{K} :\ |S| = o, \mathrm{sub}_{o-1}(\kappa) \subseteq K_{o-1}\}, \qquad \mathcal{R}_o = \{\zeta(\kappa) : \kappa \in C_o\},$$

where $\mathrm{sub}_{o-1}(\kappa)$ is the inherited-label subrule set defined above. If $C_o$ is empty, set $K_o = \emptyset$. Otherwise,

$$\tau_o = \max\{Q_{1-\rho_o}(\mathcal{R}_o), \tau_{\min}\}, \qquad K_o = \{\kappa \in C_o : \zeta(\kappa) \ge \tau_o\}.$$

The screened rule set is

$$\mathsf{HS}(\mathcal{K}, \zeta) = \bigcup_{o=1}^{q} K_o.$$

This definition is used as $K^m = \mathsf{HS}(K_q^{\mathrm{sh}}, s^m)$ and $K^{(a)} = \mathsf{HS}(K_q^{\mathrm{sh}}, s^{(a)})$. If screening produces an empty rule set, the highest-scoring first-order antecedent is retained, with ties resolved by the canonical key. The recursion uses retained lower-order subrules to gate higher-order candidates but does not assume anti-monotonic fuzzy support. Gaussian widths and firing-strength denominators are numerically lower-bounded, and unmatched transfer entries are represented by zero-row selectors.

The source and target stages use the same lower-order closure constraint, but their empirical thresholds follow the order-wise source-pool definitions in the main method. Let $K_o^{\mathrm{uni}} = \{\kappa \in K^{\mathrm{uni}} : |S| = o\}$. If $K_o^{\mathrm{uni}}$ is empty, set $K_o^{\mathrm{src}} = \emptyset$. Otherwise, for the source-side support proxy, define

$$\mathcal{R}_o^{\mathrm{src}} = \{r^{\mathrm{src}}(\kappa) : \kappa \in K_o^{\mathrm{uni}}\}, \qquad \xi_o = \max\{Q_{1-\rho_o}(\mathcal{R}_o^{\mathrm{src}}), \tau_{\min}\}.$$

Then $K_1^{\mathrm{src}} = \{\kappa \in K_1^{\mathrm{uni}} : r^{\mathrm{src}}(\kappa) \ge \xi_1\}$, and for $o \ge 2$,

$$K_o^{\mathrm{src}} = \{\kappa \in K_o^{\mathrm{uni}} : r^{\mathrm{src}}(\kappa) \ge \xi_o,\ \mathrm{sub}_{o-1}(\kappa) \subseteq K_{o-1}^{\mathrm{src}}\}.$$

The source-supported antecedent pool is $K^{\mathrm{src}} = \bigcup_{o=1}^{q} K_o^{\mathrm{src}}$. Thus, the source threshold is computed over all source rules of the same order before the lower-order closure filter is applied. If the complete source-pool screening returns $K^{\mathrm{src}} = \emptyset$, the fallback retains the first-order antecedent in $K^{\mathrm{uni}}$ with the largest $r^{\mathrm{src}}$, again using the canonical key to break ties.

For target-side rule identification, if $K_o^{\mathrm{src}}$ is empty, set $K_o^t = \emptyset$. Otherwise, define

$$\mathcal{R}_o^t = \{\hat{r}^t(\kappa) : \kappa \in K_o^{\mathrm{src}}\}, \qquad \tau_o^t = \max\{Q_{1-\rho_o}(\mathcal{R}_o^t), \tau_{\min}\}.$$

Then $K_1^t = \{\kappa \in K_1^{\mathrm{src}} : \hat{r}^t(\kappa) \geq \tau_1^t\}$, and for $o \geq 2$,

$$K_o^t = \{\kappa \in K_o^{\mathrm{src}} : \hat{r}^t(\kappa) \geq \tau_o^t,\ \mathrm{sub}_{o-1}(\kappa) \subseteq K_{o-1}^t\}.$$

The final target rule set is $K^t = \bigcup_{o=1}^{q} K_o^t$. If an order-wise candidate set or score multiset is empty, the corresponding retained set at that order is empty. If the complete target screening returns $K^t = \emptyset$, the fallback retains the first-order antecedent in $K^{\mathrm{src}}$ with the largest $\hat{r}^t$, with canonical-key tie-breaking. Since target rules are selected from $K^{\mathrm{src}}$, every final target rule has at least one matched source in $K^{\mathrm{uni}}$; sources without that rule are excluded from rule-wise fusion and contribute only through zero-padded response-level anchor quantities where explicitly defined.

An antecedent identity $\kappa = (S, l_S)$ is a cross-center matching key rather than a complete TSK rule because it has no consequent. At center $u$, $\kappa$ and the center-specific consequent $\mathbf{p}_\kappa^u$ define the zero-order rule

$$\mathcal{R}_\kappa^u : \quad \text{IF } x_j \text{ is } A_{j,l_j}^m,\ j \in S, \quad \text{THEN } \mathbf{p}_\kappa^u,$$

and features outside $S$ do not participate in that antecedent. In the auxiliary objective, the matched-consequent term aligns $\mathbf{P}_{\mathrm{tr}}^{(a)}$ with $\mathbf{P}_{\mathrm{tr}}^m$, while the response term aligns their average rule-activation-weighted responses. In the target stage, $\widetilde{\mathbf{\Gamma}}_{K^t}^u$ is computed from the source design matrix on its own retained rule set and then reordered and zero-padded to $K^t$, without renormalizing firing strengths after alignment. The implemented $d_u$ is an unnormalized squared support-profile discrepancy over $K^t$; it is used as a heuristic compatibility score rather than a calibrated measure of transferability. Its scale depends on the number and dispersion of retained-rule differences. The experiments fix $\tau_\omega = 0.1$ globally; no rule-count normalization or temperature-sensitivity analysis is reported.

### B.2 Algorithmic Summary and Implementation Conventions

Table S2 summarizes the training sequence needed for reproduction. Antecedent identities are stable keys $(S, l_S)$, not center-specific row indices; unions, intersections, presence indicators, and zero-filled alignment are therefore computed in the shared identity space induced by $\Psi^m$. Higher-order candidates are gated by retained lower-order subrules as a rule-growth constraint, without assuming anti-monotonic fuzzy support.

Table S2: Algorithmic summary of MCRT-TSK under the shared rule-identity protocol.

| Item | Description |
|---|---|
| Input | Main-center data $\mathcal{D}^m$, auxiliary-center data $\{\mathcal{D}^{(a)}\}_{a=1}^{A}$, target-center data $\mathcal{D}^t$, maximum rule order $q$, maximum one-dimensional fuzzy granularity $L_{\max}$, feature-granularity penalty $\lambda_L$, retention ratios $\{\rho_r\}_{r=1}^{q}$, threshold floor $\tau_{\min}$, consequent regularization parameters $\lambda^m, \lambda_1^{(a)}, \lambda_2^{(a)}, \lambda_3^{(a)}, \lambda_1^t, \lambda_2^t, \lambda_3^t$ with $\lambda_2^{(a)} > 0$, target-balance constant $n_0$, fusion temperature $\tau_\omega$, and optional source weights $\{\pi_u\}$. If $\{\pi_u\}$ is not provided, normalized source sample counts are used. |
| 1 | For each feature $j = 1, \ldots, d$, enumerate $L = 1, \ldots, L_{\max}$, evaluate the deterministic one-dimensional clustering routine on the main-center feature values, and choose $L_j$ by the penalized distortion criterion. Store the ordered Gaussian centers and widths as the shared antecedent partition $\Psi^m$. |
| 2 | Generate $K_q^{\mathrm{sh}} = \{(S, l_S) : 1 \leq \lvert S\rvert \leq q,\ l_j \in \{1, \ldots, L_j\}\}$. Assign each rule a stable key determined by its feature subset $S$ and fuzzy-set index tuple $l_S$, and sort rules by order, feature indices, and fuzzy-set indices. |
| 3 | Compute firing strengths on $\mathcal{D}^m$ for all rules in $K_q^{\mathrm{sh}}$. For each order $r = 1, \ldots, q$, keep only candidates whose $(r-1)$-order subrules have already been retained, compute main support $s^m(\kappa)$ on the closed candidate set, set $\tau_r^m = \max\{Q_{1-\rho_r}(s^m), \tau_{\min}\}$, and retain $K_r^m = \{\kappa : s^m(\kappa) \geq \tau_r^m\}$. Set $K^m = \bigcup_{r=1}^{q} K_r^m$. If $K^m = \emptyset$, retain the top-scoring first-order antecedent with canonical-key tie-breaking. |
| 4 | Build the normalized design matrix $\mathbf{\Phi}^m$ on $K^m$, solve the ridge system for $\mathbf{P}^m$, and set $\mathbf{\Gamma}^m = N_m^{-1}\mathbf{1}^\top \mathbf{\Phi}^m$. The main source model summary is $\mathcal{S}^m = (K^m, \mathbf{P}^m, s^m, \mathbf{\Gamma}^m, N_m)$, with $\Psi^m$ shared globally. |

Table S2 (continued)

| Item | Description |
|---|---|
| 5 | For each auxiliary center $a$, reuse $\Psi^m$ and $K_q^{\mathrm{sh}}$, repeat the same order-wise support screening to obtain $K^{(a)}$, applying the same top-scoring first-order fallback if the complete auxiliary retained set is empty, match $K^{(a)} \cap K^m$ by rule keys, and solve the auxiliary consequent problem with local supervision, a main-consequent anchor, mean-response alignment over the matched rules, and ridge regularization. Store $\mathcal{S}^{(a)} = (K^{(a)}, \mathbf{P}^{(a)}, s^{(a)}, \mathbf{\Gamma}^{(a)}, N_a)$, where $\mathbf{\Gamma}^{(a)} = N_a^{-1}\mathbf{1}^\top \mathbf{\Phi}^{(a)}$. |
| 6 | Let $\mathcal{U} = \{m, 1, \ldots, A\}$ denote the source index set with source models $\{\mathcal{S}^m, \mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(A)}\}$. Construct $K^{\mathrm{uni}}$ as the key-based union of all source rules. Align every source model to $K^{\mathrm{uni}}$ with zero-filled rows/columns for missing rules. |
| 7 | For each $\kappa \in K^{\mathrm{uni}}$, compute the weighted source support $r^{\mathrm{src}}(\kappa) = \sum_{u\in\mathcal{U}} \pi_u \chi_u(\kappa) s^u(\kappa)$, where $\chi_u(\kappa)$ indicates whether source $u$ retains the rule and missing rules contribute zero. |
| 8 | For each order in $K^{\mathrm{uni}}$, compute the source threshold from all rules of that order using $r^{\mathrm{src}}$, then apply the order-wise closure constraint and retain rules whose $r^{\mathrm{src}}$ values meet or exceed the threshold. The retained source-supported antecedent pool is $K^{\mathrm{src}}$. If it is empty, keep the first-order antecedent with the largest $r^{\mathrm{src}}$. |
| 9 | Compute target support $s^t(\kappa)$ for $\kappa \in K^{\mathrm{src}}$. Set $\beta_t = N_t/(N_t + n_0)$ and $\hat{r}^t(\kappa) = \beta_t s^t(\kappa) + (1-\beta_t) r^{\mathrm{src}}(\kappa)$. For each order in $K^{\mathrm{src}}$, compute the threshold from all rules of that order using $\hat{r}^t$, then apply order-wise closed screening to obtain $K^t \subseteq K^{\mathrm{src}}$. If $K^t = \emptyset$, keep the first-order antecedent with the largest $\hat{r}^t$. |
| 10 | Align each source model to $K^t$ by rule identity. Compute each zero-padded source profile distance $d_u = \sum_{\kappa\in K^t} (\chi_u(\kappa) s^u(\kappa) - \hat{r}^t(\kappa))^2$. Use $\exp(-d_u/\tau_\omega)$ to obtain global source weights $\alpha_u$ and rule-wise weights $\omega_{u,\kappa}$ over sources that contain $\kappa$. Construct the fused consequent prior $\mathbf{P}^t_{\mathrm{fus}}$ rule by rule and the multi-source response anchor $\mathbf{r}^t_{\mathrm{fus}} = \sum_{u\in\mathcal{U}} \alpha_u \widetilde{\mathbf{\Gamma}}^u_{K^t} \widetilde{\mathbf{P}}^u_{K^t}$, where $\widetilde{\mathbf{\Gamma}}^u_{K^t}$ is the zero-padded, $K^t$-aligned source average activation vector without post-alignment renormalization. |
| 11 | Build $\mathbf{\Phi}^t$ on $K^t$, set $\mathbf{\Gamma}^t = N_t^{-1}\mathbf{1}^\top \mathbf{\Phi}^t$, and solve $A_t \mathbf{P}^t = B_t$, where $A_t = (\mathbf{\Phi}^t)^\top \mathbf{\Phi}^t + (\lambda_1^t + \lambda_2^t)\mathbf{I} + \lambda_3^t (\mathbf{\Gamma}^t)^\top \mathbf{\Gamma}^t$ and $B_t = (\mathbf{\Phi}^t)^\top \mathbf{Y}^t + \lambda_1^t \mathbf{P}^t_{\mathrm{fus}} + \lambda_3^t (\mathbf{\Gamma}^t)^\top \mathbf{r}^t_{\mathrm{fus}}$. |
| Output | Target-center model $\mathcal{M}^t = \{\Psi^m, K^t, \mathbf{P}^t\}$. For a new target sample $\mathbf{x}$, compute its normalized firing vector on $K^t$, class-score vector $\mathbf{f}^t(\mathbf{x}) = (\boldsymbol{\phi}^t(\mathbf{x}))^\top \mathbf{P}^t$, and prediction $\hat{y} = \arg\max_c f^t_c(\mathbf{x})$. The scores are not treated as calibrated probabilities. |

### B.3 Interpretation of the Weighting Components

The weighting quantities have operational rather than probabilistic interpretations. The score $r^{\mathrm{src}}$ summarizes sample-count-weighted source retention and activation support rather than predictive transferability. The factor $\beta_t = N_t/(N_t + n_0)$ depends only on the target sample count and does not encode task uncertainty, dimensionality, class count, or imbalance. The global source discrepancy $d_u$ does not include a rule-specific held-out compatibility loss. Moreover, the main-center partition defines the available antecedent dictionary, so target-only regions cannot introduce new identities. These components are implementation-matched heuristics rather than calibrated or scale-invariant estimators.

## C Supplementary Experimental Details

### C.1 Dataset Details

The benchmark contains 18 UCI tabular classification datasets (Kelly et al., 2026): 10 binary and 8 multi-class tasks spanning 150–5473 samples, 4–60 features, and 2–7 classes (Table S3). Before center construction, each dataset is represented as a numerical feature matrix with labels encoded as $1, \ldots, C$. The four centers are simulated subsets of a single processed dataset rather than independently acquired domains. Table S3 reports the first-level $M/A_1/A_2/T$ pool sizes before the center-wise 8:2 split; abbreviations are used consistently throughout the paper and supplementary material.

Table S3: Dataset details and abbreviation map used in the main-text analysis. All displayed datasets are from the UCI Machine Learning Repository. Here $N$, $d$, and $C$ denote the number of samples, features, and classes, respectively. The center-size column reports first-level sample counts as $M/A_1/A_2/T$, where $M$, $A_1$, $A_2$, and $T$ denote the main center, the two auxiliary centers, and the target center before the center-wise 8:2 train/test split. Rows are ordered by task type (binary before multi-class) and then by increasing $N$.

| Abbr. | Full dataset name | $N$ | $d$ | $C$ | Center $N$ ($M/A_1/A_2/T$) | Task |
|---|---|---|---|---|---|---|
| Son | Sonar (Mines vs. Rocks) | 208 | 60 | 2 | 62/52/52/42 | Binary |
| Hea | Heart Disease (Cleveland; complete cases) | 297 | 13 | 2 | 89/74/74/60 | Binary |
| Bre | Breast Cancer Wisconsin (Diagnostic) | 569 | 30 | 2 | 171/142/142/114 | Binary |
| Aus | Australian Credit Approval | 690 | 14 | 2 | 207/173/173/137 | Binary |
| Blo | Blood Transfusion Service Center | 748 | 4 | 2 | 224/188/186/150 | Binary |
| Dia | Pima Indians Diabetes | 768 | 8 | 2 | 230/192/192/154 | Binary |
| Qsa | QSAR biodegradation | 1055 | 41 | 2 | 316/264/264/211 | Binary |
| DiR | Diabetic retinopathy Debrecen | 1151 | 19 | 2 | 345/288/288/230 | Binary |
| Ban | Banknote authentication | 1372 | 4 | 2 | 412/344/342/274 | Binary |
| Spa | Spambase | 4601 | 57 | 2 | 1380/1150/1150/921 | Binary |
| Iri | Iris | 150 | 4 | 3 | 45/39/36/30 | Multi-class |
| Win | Wine | 178 | 13 | 3 | 52/45/45/36 | Multi-class |
| Con | Contraceptive Method Choice | 1473 | 9 | 3 | 442/368/368/295 | Multi-class |
| Car | Car evaluation | 1728 | 6 | 4 | 519/432/431/346 | Multi-class |
| WiQ | Wine Quality (white; labels 5–7) | 4535 | 11 | 3 | 1360/1134/1133/908 | Multi-class |
| WiD | Wine Quality (white; seven observed labels) | 4898 | 11 | 7 | 1469/1225/1224/980 | Multi-class |
| Wav | Waveform Database Generator (Version 1) | 5000 | 21 | 3 | 1500/1250/1250/1000 | Multi-class |
| Pag | Page Blocks Classification | 5473 | 10 | 5 | 1641/1368/1368/1096 | Multi-class |

### C.2 Parameter Settings for Compared Methods

The supplementary material reports the parameterization of all compared methods. Table S4 summarizes the hyperparameter spaces and fixed configurations used in the aligned MATLAB experiment package. Whenever the released implementation permits, all methods use a common outer protocol: the same processed datasets and label encoding; one simulated main center, two simulated auxiliary centers, and one simulated target center with 30%/25%/25%/20% first-level proportions; label stratification when feasible during center construction and each center-wise 8:2 train/test split; ten repetitions with seeds 2–11; and target-center test reporting. No feature transform, class-prior intervention, labeling-rule change, or center-specific sampling mechanism is introduced to create covariate shift, label shift, or concept shift. Method-specific hyperparameters are selected by five-fold validation when the implementation exposes a tunable search space.

Table S4: Hyperparameter spaces and fixed configurations for MCRT-TSK and the compared algorithms under the aligned experiment package. $K$ denotes the number of fuzzy rules where applicable. BC, AC, and TC denote the base (main-center), auxiliary-center, and target-center stages, respectively.

| Method | Hyperparameter space or fixed configuration |
|---|---|
| MCRT-TSK (ours) | The main comparison uses $q = 2$. The $q$-sensitivity study additionally evaluates $q \in \{1, 2, 3\}$. The remaining settings are $L_{\max} = 3$, $\boldsymbol{\rho} = (0.8, 0.9)$, $\tau_{\min} \in \{10^{-4}, 10^{-3}\}$, $\eta \in \{0.5, 1, 2\}$, the feature-granularity penalty $\lambda_L \in \{10, 1, 0.01\}$, $n_0 = 20$, and $\tau_\omega = 0.1$; the consequent regularization parameters in the main, auxiliary, and target objectives are selected from $\{10, 1, 0.01\}$. |
| 0-TSK and 1-TSK | Zero-order and first-order TSK models are evaluated. The shared grid is $K \in \{5, 10, \ldots, 30\}$, kernel width $\in \{0.01, 0.1, 1, 10, 100\}$, ridge parameter $\lambda \in \{0.01, 0.1, 1, 10, 100\}$, $\tau \in \{0.01, 0.1, 1, 10, 100\}$, and GLS loss. |
| TL-SSL-TSK | $m \in \{1.1, 1.5, 2\}$, $K \in \{5, 10, \ldots, 30\}$, $\lambda_1, \lambda_2, \lambda_3 \in \{0.01, 0.1, 1, 10\}$, and $\eta \in \{0, 0.2, \ldots, 1\}$. |
| ETTL-TSK-FS | $K \in \{5, 10, \ldots, 30\}$, $\lambda_1 \in \{0.01, 0.1, 1, 10\}$, and $\tau \in \{0.01, 0.1, 1, 10\}$. |

Table S4 (continued)

| Method | Hyperparameter space or fixed configuration |
|---|---|
| FL-TSK-FRBSs | Fixed $K = 30$, with the fuzzy input domain set to $[0, 1]$. |
| MKTC-R0T | BC: $K \in \{5, 10, \ldots, 30\}$, $\lambda_1, \lambda_2 \in \{0.1, 1, 10\}$. AC: $\lambda_1 \in \{0.1, 1, 10\}$, $\lambda_2 \in \{0.01, 0.1, 1, 10\}$. For target integration, the aligned wrapper and invoked integration routine use $\gamma \in \{0.01, 0.1, 1, 10\}$. |
| MIP-TRL-FS | Budgeted 48-configuration subset: $K \in \{3, 5, 7\}$, subspace dimension $\in \{10, 20\}$, maximum iterations $= 10$, fuzzy index $= 2$, $\alpha, \beta \in \{0.1, 1\}$, $\lambda \in \{0.01, 0.1\}$, neighbors $= 5$, views $= \{\texttt{raw}, \texttt{signedsqrt}\}$, PCA dimension $= 16$, and $\epsilon = 10^{-8}$. |
| MTSK-DA | BC: $K \in \{5, 10, \ldots, 30\}$, kernel width and $\tau \in \{0.01, 0.1, 1, 10, 100\}$, $\lambda_1, \lambda_2 \in \{0.1, 1, 10\}$. AC: $\lambda_1, \lambda_2, \lambda_3 \in \{0.1, 1, 10\}$, kernel width $\in \{0.01, 0.1, 1, 10, 100\}$. TC: kernel width $\in \{0.01, 0.1, 1, 10, 100\}$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \delta \in \{0.1, 1, 10\}$, $\beta = 1$, maximum iterations $= 30$. |
| KCF-LSG-MTSK | BC uses the same $K$, kernel-width, $\tau$, $\lambda_1$, and $\lambda_2$ ranges as MTSK-DA. AC: $\lambda_1, \lambda_2, \lambda_3 \in \{0.1, 1, 10\}$, kernel width $\in \{0.1, 1, 10\}$. TC: $\lambda_1, \lambda_2, \lambda_3 \in \{0.1, 1, 10\}$, $\lambda_4 \in \{0.001, 0.01, 0.1\}$, $\delta \in \{0.1, 1, 10\}$, $\beta = 1$, kernel width $\in \{0.01, 0.1, 1, 10, 100\}$. |

In panel (a) of main-text Fig. 3(B), the error bars are normal-approximation 95% confidence intervals across the 15 dataset-level ten-run means, computed as $\bar{x} \pm 1.96s/\sqrt{15}$. They are neither intervals over the ten repetitions nor $t$-based or bootstrap intervals. Panel (d) of Fig. 3(B) reports dataset-level paired ACC differences without confidence intervals. The unavailable $q = 3$ results arise from resource failures rather than performance-based exclusions.

The parameter ranges reported for MIP-TRL-FS (Li et al., 2024a) define a 36015-configuration Cartesian grid. The comparison uses the disclosed 48-configuration subset in Table S4, preserving all rule-count choices, 10 iterations, the reported default configuration ($K = 3$, $d' = 10$, fuzzy index $= 2$, $\alpha = \beta = 1$, $\lambda = 0.01$), and nearby low-to-moderate values. This is a budgeted comparison rather than exhaustive tuning, so the reported MIP-TRL-FS results are conditional on that search budget.

### C.3 Additional Performance Metrics

For a $C$-class test set, let $n$ denote its size, and let $\mathrm{TP}_c$, $\mathrm{FP}_c$, and $\mathrm{FN}_c$ denote the one-vs-rest counts for class $c$. A class-wise precision or recall is set to zero when its denominator is zero. The reported metrics are

$$\mathrm{ACC} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\hat{y}_i = y_i), \quad \text{Macro-F1} = \frac{1}{C}\sum_{c=1}^{C} \frac{2P_cR_c}{P_c + R_c}, \quad \mathrm{GM} = \left(\prod_{c=1}^{C} R_c\right)^{1/C},$$

where $P_c = \mathrm{TP}_c/(\mathrm{TP}_c + \mathrm{FP}_c)$ and $R_c = \mathrm{TP}_c/(\mathrm{TP}_c + \mathrm{FN}_c)$, with the F1 term set to zero when $P_c + R_c = 0$. ACC measures overall correctness, Macro-F1 weights classes equally, and G-mean is zero when any class has zero recall, making class collapse visible.

Tables S5 and S6 report target-test Macro-F1 and G-mean (GM), respectively, over the same ten repetitions and under the configurations selected by ACC; no Macro-F1- or GM-specific re-ranking is performed. UMC denotes an unsupported multi-class setting in the released implementation, TO denotes a timeout ($> 120$ hours per run), and "–" denotes an unavailable metric output despite a valid ACC result.

Under the ACC-selected configurations, MCRT-TSK attains the highest Macro-F1 on 11 of the 18 datasets and the highest or tied-highest G-mean on 9. These results provide complementary class-wise evidence; differences from the ACC rankings are expected because hyperparameters were not re-selected for either metric.

Table S5: Datasets Son–Spa are binary; Iri–Pag are multi-class. Performance comparison of MCRT-TSK and nine comparators on 18 datasets (Macro-F1↑, %; mean (standard deviation)). Boldface and underlining indicate the best and second-best valid results, respectively. MCRT-TSK reports Macro-F1 values under the ACC-selected configurations; no metric-specific re-ranking is applied. UMC: unsupported multi-class; TO: timeout ($>$ 120 hours per run); –: metric output unavailable despite a valid ACC result. Means are shown above their parenthesized standard deviations.

| | (a) Target-only diagnostics and transfer comparators | | | | | (b) Remaining transfer comparators and MCRT-TSK | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | 0-TSK | 1-TSK | TL-SSL -TSK | ETTL -TSK-FS | FL-TSK -FRBSs | MKTC -R0T | MIP -TRL-FS | MTSK -DA | KCF-LSG -MTSK | MCRT-TSK (ours) |
| Son | 48.94 (21.10) | 47.18 (19.61) | TO | 75.22 (21.74) | 47.18 (15.88) | 62.56 (15.03) | 62.74 (16.06) | – | 79.13 (13.25) | **84.13 (12.29)** |
| Hea | 78.09 (9.82) | 49.62 (19.03) | 76.56 (12.79) | 78.11 (11.13) | 48.58 (23.22) | 73.10 (12.86) | 73.91 (14.96) | 79.64 (10.68) | 80.54 (10.60) | **88.18 (7.14)** |
| Bre | 92.61 (2.54) | 51.93 (20.37) | 94.97 (4.73) | 94.01 (5.22) | 70.36 (8.13) | 93.13 (3.38) | 76.48 (14.79) | 94.97 (2.32) | 94.62 (4.70) | **96.42 (3.49)** |
| Aus | 82.60 (8.71) | 71.22 (14.94) | 86.42 (5.11) | 88.13 (6.61) | 35.71 (0.00) | 82.41 (6.59) | 71.32 (10.24) | 89.14 (6.68) | 84.87 (8.25) | **90.62 (3.68)** |
| Blo | 54.92 (13.17) | 28.97 (15.81) | 55.40 (14.10) | 59.77 (10.57) | 59.75 (12.95) | 54.41 (10.86) | 52.52 (12.52) | 59.38 (7.48) | – | **65.34 (11.85)** |
| Dia | 63.37 (9.25) | 56.56 (9.65) | 66.19 (8.21) | 66.74 (7.91) | 67.75 (7.61) | 66.77 (9.69) | 50.92 (9.83) | 62.60 (6.12) | **69.76 (10.93)** | 67.78 (7.05) |
| Qsa | 76.86 (7.31) | 37.87 (15.57) | **80.71 (8.44)** | 79.34 (5.88) | 76.70 (5.12) | 69.68 (11.92) | 68.52 (11.24) | 79.49 (6.94) | 78.39 (7.81) | 77.99 (3.24) |
| DiR | 60.93 (5.56) | 48.39 (12.58) | 63.59 (6.02) | 63.10 (6.59) | 61.75 (8.62) | 59.54 (9.23) | 53.39 (6.10) | 61.52 (6.51) | 63.70 (5.84) | **64.92 (2.46)** |
| Ban | 98.70 (1.77) | **99.44 (1.26)** | 96.83 (1.77) | 96.64 (2.61) | 96.83 (2.31) | 98.69 (1.26) | 70.43 (16.95) | 98.49 (1.96) | 98.31 (2.25) | 95.86 (3.30) |
| Spa | 83.62 (4.49) | 73.03 (18.59) | 87.32 (2.87) | 82.95 (2.65) | **89.73 (1.72)** | 87.56 (3.15) | 80.37 (3.37) | TO | TO | 85.63 (1.27) |
| Iri | 90.22 (14.79) | 65.57 (26.27) | UMC | UMC | UMC | UMC | 66.33 (37.06) | – | – | **100.00 (0.00)** |
| Win | 92.78 (12.43) | 48.29 (24.49) | UMC | UMC | UMC | UMC | 50.37 (32.29) | 95.78 (9.63) | 97.33 (5.62) | **100.00 (0.00)** |
| Con | 43.11 (7.54) | 46.57 (8.76) | UMC | UMC | UMC | UMC | 32.65 (5.48) | **50.57 (5.02)** | 48.22 (7.91) | 42.50 (5.73) |
| Car | 34.22 (4.91) | **45.73 (9.23)** | UMC | UMC | UMC | UMC | 26.94 (10.51) | – | – | 39.64 (8.96) |
| WiQ | 50.27 (4.06) | 52.36 (3.19) | UMC | UMC | UMC | UMC | 48.44 (5.22) | TO | 53.21 (4.68) | **53.35 (3.35)** |
| WiD | 20.60 (1.82) | 22.86 (2.24) | UMC | UMC | UMC | UMC | 24.51 (3.98) | TO | – | **25.58 (2.89)** |
| Wav | 81.33 (1.74) | 80.38 (2.04) | UMC | UMC | UMC | UMC | 74.69 (3.50) | TO | 82.60 (3.39) | **84.29 (1.09)** |
| Pag | 43.69 (10.20) | **57.55 (6.73)** | UMC | UMC | UMC | UMC | 50.83 (13.01) | TO | TO | 54.87 (9.64) |

Table S6: Datasets Son–Spa are binary; Iri–Pag are multi-class. Performance comparison of MCRT-TSK and nine comparators on 18 datasets (G-mean↑, %; mean (standard deviation)). Boldface and underlining indicate the best and second-best valid results, respectively. MCRT-TSK reports G-mean values under the ACC-selected configurations; no metric-specific re-ranking is applied. UMC: unsupported multi-class; TO: timeout ($>$ 120 hours per run); –: metric output unavailable despite a valid ACC result. Means are shown above their parenthesized standard deviations.

| | (a) Target-only diagnostics and transfer comparators | | | | | (b) Remaining transfer comparators and MCRT-TSK | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Data | 0-TSK | 1-TSK | TL-SSL -TSK | ETTL -TSK-FS | FL-TSK -FRBSs | MKTC -R0T | MIP -TRL-FS | MTSK -DA | KCF-LSG -MTSK | MCRT-TSK (ours) |
| Son | 39.05 (30.73) | 40.96 (26.56) | TO | 73.97 (22.75) | 35.48 (27.46) | 61.09 (15.13) | 61.41 (15.93) | 58.86 (27.09) | 77.87 (14.39) | **82.68 (13.53)** |
| Hea | 77.61 (9.84) | 38.11 (30.79) | 75.25 (13.77) | 76.71 (12.62) | 27.83 (38.13) | 71.68 (13.65) | 73.39 (14.82) | 78.92 (10.79) | 79.97 (10.80) | **87.75 (7.42)** |
| Bre | 92.26 (2.68) | 52.54 (19.63) | 94.13 (5.11) | 93.66 (5.91) | 67.93 (7.86) | 93.01 (3.23) | 75.45 (15.88) | 93.85 (2.41) | 94.31 (4.30) | **95.43 (4.46)** |
| Aus | 82.29 (9.00) | 67.25 (25.08) | 86.20 (5.26) | 87.60 (7.34) | 0.00 (0.00) | 82.49 (6.70) | 70.36 (10.19) | 89.19 (6.70) | 84.59 (8.57) | **90.69 (3.99)** |
| Blo | 32.28 (29.29) | 15.56 (26.92) | 30.09 (28.68) | 49.82 (16.68) | 41.04 (26.24) | 34.85 (27.49) | 41.81 (24.16) | **56.61 (11.17)** | 33.61 (25.76) | 49.95 (22.04) |
| Dia | 59.99 (10.58) | 53.99 (9.48) | 63.19 (11.60) | 64.12 (9.91) | 64.15 (9.13) | 65.93 (13.61) | 41.24 (18.92) | 60.79 (7.83) | **68.56 (13.49)** | 61.90 (11.00) |
| Qsa | 75.97 (7.25) | 25.95 (28.42) | 80.30 (8.67) | 77.27 (6.63) | 78.44 (4.58) | 65.76 (24.21) | 68.07 (12.91) | **80.68 (6.41)** | 79.34 (7.55) | 72.46 (4.36) |
| DiR | 60.56 (5.50) | 40.78 (23.53) | 63.54 (6.07) | 62.97 (6.67) | 61.35 (9.03) | 58.63 (10.04) | 52.23 (6.43) | 61.32 (6.44) | 63.52 (6.06) | **64.65 (2.69)** |
| Ban | 98.81 (1.61) | **99.45 (1.20)** | 97.01 (1.70) | 96.76 (2.52) | 96.94 (2.12) | 98.74 (1.19) | 67.44 (19.43) | 98.38 (2.20) | 98.32 (2.26) | 95.84 (3.48) |
| Spa | 82.83 (5.64) | 73.14 (19.21) | 86.14 (3.40) | 83.00 (3.08) | **88.91 (2.20)** | 87.68 (2.96) | 79.67 (3.65) | TO | TO | 82.76 (1.51) |
| Iri | 83.81 (31.01) | 42.17 (45.68) | UMC | UMC | UMC | UMC | 60.11 (43.12) | 83.81 (31.01) | 83.81 (31.01) | **100.00 (0.00)** |
| Win | 92.60 (12.88) | 29.54 (38.48) | UMC | UMC | UMC | UMC | 41.38 (38.11) | 95.67 (10.07) | 97.47 (5.33) | **100.00 (0.00)** |
| Con | 38.34 (9.76) | 38.06 (21.27) | UMC | UMC | UMC | UMC | 23.48 (16.53) | **49.17 (5.67)** | 46.51 (9.31) | 16.55 (21.85) |
| Car | 0.00 (0.00) | **9.83 (20.75)** | UMC | UMC | UMC | UMC | 3.99 (12.61) | 0.00 (0.00) | 4.94 (15.63) | 0.00 (0.00) |
| WiQ | 43.81 (6.85) | 48.53 (3.50) | UMC | UMC | UMC | UMC | 46.52 (6.41) | TO | **49.18 (5.02)** | 46.92 (5.11) |
| WiD | **0.00 (0.00)** | **0.00 (0.00)** | UMC | UMC | UMC | UMC | **0.00 (0.00)** | TO | **0.00 (0.00)** | **0.00 (0.00)** |
| Wav | 81.14 (1.75) | 80.28 (2.00) | UMC | UMC | UMC | UMC | 74.39 (3.83) | TO | 82.12 (3.68) | **83.87 (1.16)** |
| Pag | 0.00 (0.00) | 12.39 (26.16) | UMC | UMC | UMC | UMC | **23.36 (30.40)** | TO | TO | 5.71 (18.07) |

### C.4 Statistical Analysis

Tables S7–S9 provide the numerical evidence behind main-text Fig. 2(A–B). Every rank, paired margin, test statistic, correction order, and decision can be traced to the valid ACC entries used by the corresponding test.

The global Friedman test and the pairwise Wilcoxon tests answer related but different questions. Friedman ranking requires all compared methods to be evaluated on exactly the same datasets. Therefore, Table S7 uses only the strict complete all-method subset. Since TL-SSL-TSK, ETTL-TSK-FS, FL-TSK-FRBSs, and MKTC-R0T do not provide valid multi-class outputs in the benchmark matrix, the strict complete subset contains only eight binary datasets. The resulting Friedman ranks should thus be read as common binary-dataset evidence across all ten methods, not as a mixed binary–multi-class omnibus comparison.

The paired Wilcoxon analysis in Table S8 uses a different eligibility rule. For each baseline, MCRT-TSK is compared only on datasets for which both methods have valid ACC values. This rule preserves multi-class evidence when a baseline supports it, while excluding unsupported multi-class cases (UMC) and timeout entries (TO) from the paired test. The B/M column reports the binary and multi-class composition of the valid paired set; the UMC and TO columns report how many entries were excluded by status rather than converted into numerical losses. The mean and median $\Delta$ACC values measure the magnitude of the paired advantage, and W/T/L reports how often MCRT-TSK wins, ties, or loses on the valid paired datasets.

Table S9 further separates uncorrected paired evidence from the multiple-comparison decision. The raw Wilcoxon $p$-value indicates whether a paired margin is unlikely under the signed-rank null for a single baseline comparison, whereas the Holm-adjusted decision controls the family-wise error rate across the nine baseline comparisons. Under this corrected criterion, seven comparisons remain significant. The two non-significant comparisons, TL-SSL-TSK and FL-TSK-FRBSs, both have positive paired ACC margins but are based on binary-only paired evidence and do not pass Holm correction.

Table S7: Friedman ranks on the strict complete ACC subset. The subset contains eight datasets (Hea, Bre, Aus, Blo, Dia, Qsa, DiR, and Ban) and excludes status-only entries rather than imputing them.

| Method | Avg. rank | Datasets |
|---|---|---|
| MCRT-TSK (ours) | 2.06 | 8 |
| TL-SSL-TSK | 3.88 | 8 |
| KCF-LSG-MTSK | 4.00 | 8 |
| ETTL-TSK-FS | 4.56 | 8 |
| MTSK-DA | 4.69 | 8 |
| 0-TSK | 5.56 | 8 |
| MKTC-R0T | 6.38 | 8 |
| FL-TSK-FRBSs | 6.44 | 8 |
| 1-TSK | 8.50 | 8 |
| MIP-TRL-FS | 8.94 | 8 |

Table S8: Paired evidence for Wilcoxon tests. $n$ is the number of valid paired datasets; B/M reports the binary and multi-class split. UMC and TO denote unsupported multi-class and timeout ($> 120$ hours/run) entries excluded from the test. $\Delta$ACC is MCRT-TSK minus the baseline in percentage points.

| Method | $n$ | B/M | UMC | TO | Mean $\Delta$ | Median $\Delta$ | W/T/L |
|---|---|---|---|---|---|---|---|
| 0-TSK | 18 | 10/8 | 0 | 0 | +6.18 | +4.17 | 17/0/1 |
| 1-TSK | 18 | 10/8 | 0 | 0 | +18.98 | +13.43 | 17/0/1 |
| TL-SSL-TSK | 9 | 9/0 | 8 | 1 | +2.56 | +1.52 | 6/1/2 |
| ETTL-TSK-FS | 10 | 10/0 | 8 | 0 | +3.81 | +2.87 | 9/0/1 |
| FL-TSK-FRBSs | 10 | 10/0 | 8 | 0 | +12.70 | +3.90 | 8/0/2 |
| MKTC-R0T | 10 | 10/0 | 8 | 0 | +6.78 | +6.06 | 8/0/2 |
| MIP-TRL-FS | 18 | 10/8 | 0 | 0 | +16.80 | +14.86 | 18/0/0 |
| MTSK-DA | 13 | 9/4 | 0 | 5 | +6.60 | +4.29 | 12/0/1 |
| KCF-LSG-MTSK | 16 | 9/7 | 0 | 2 | +4.97 | +2.86 | 15/0/1 |

Table S9: Compact Holm-adjusted Wilcoxon–Friedman summary. Holm adj. $p$ is the Holm-adjusted paired Wilcoxon $p$-value against MCRT-TSK and is evaluated against 0.05; $\alpha_i$ reports the corresponding step-down raw-$p$ threshold. The Friedman $p$-value and decision are global for the strict complete all-method subset and are shown once.

| | Friedman | Unadjusted paired Wilcoxon | | | Holm-adjusted paired Wilcoxon | | |
|---|---|---|---|---|---|---|---|
| Method | rank | $z$ | $p$ | $H_0$ | adj. $p$ | $\alpha_i$ | $H_0$ |
| 0-TSK | 5.56 | 3.593 | $3.27 \times 10^{-4}$ | **Reject** | 0.0026 | 0.0063 | **Reject** |
| 1-TSK | 8.50 | 3.462 | $5.36 \times 10^{-4}$ | **Reject** | 0.0037 | 0.0071 | **Reject** |
| TL-SSL-TSK | 3.88 | 2.100 | 0.0357 | **Reject** | 0.0568 | 0.0500 | Not rej. |
| ETTL-TSK-FS | 4.56 | 2.599 | 0.0093 | **Reject** | 0.0374 | 0.0125 | **Reject** |
| FL-TSK-FRBSs | 6.44 | 2.191 | 0.0284 | **Reject** | 0.0568 | 0.0250 | Not rej. |
| MKTC-R0T | 6.38 | 2.497 | 0.0125 | **Reject** | 0.0375 | 0.0167 | **Reject** |
| MIP-TRL-FS | 8.94 | 3.724 | $1.96 \times 10^{-4}$ | **Reject** | 0.0018 | 0.0056 | **Reject** |
| MTSK-DA | 4.69 | 2.831 | 0.0046 | **Reject** | 0.0232 | 0.0100 | **Reject** |
| KCF-LSG-MTSK | 4.00 | 3.206 | 0.0013 | **Reject** | 0.0081 | 0.0083 | **Reject** |
| MCRT-TSK (ours) | 2.06 | – | – | – | – | – | – |

Friedman global test: $p = 4.50 \times 10^{-5}$; **Reject** $H_0$.

### C.5 Ablation Study

Table S10 defines the ablations shown in main-text Fig. 2(C–D). A $\surd$ marks an active component, whereas a $\times$ marks a disabled module, regularizer, or source–target scoring term. The table separates source-pool filtering $K^{\mathrm{src}}$, target-rule filtering $K^t$, and mixed source–target scoring for $K^t$.

Table S10: Ablation variants and enabled or disabled MCRT-TSK modules. The symbol $\surd$ means that a module is enabled; $\times$ means that the corresponding module or source–target scoring term is disabled.

**(a) Enabled components**

| Variant | Aux | AuxTr | $K^{\mathrm{src}}$ | $K^t$ | Src+Tgt | $\mathbf{P}^t_{\mathrm{fus}}$ | $\mathbf{r}^t_{\mathrm{fus}}$ |
|---|---|---|---|---|---|---|---|
| Full MCRT-TSK | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| w/o aux | $\times$ | $\times$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| w/o aux transfer | $\surd$ | $\times$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| w/o $K^{\mathrm{src}}$ | $\surd$ | $\surd$ | $\times$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| w/o $K^t$ | $\surd$ | $\surd$ | $\surd$ | $\times$ | $\times$ | $\surd$ | $\surd$ |
| target-only $K^t$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\times$ | $\surd$ | $\surd$ |
| w/o $p$-fusion | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\times$ | $\surd$ |
| w/o $r$-fusion | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\times$ |

**(b) Exact changes**

| Variant | Changed setting |
|---|---|
| Full MCRT-TSK | Complete MCRT-TSK pipeline. |
| w/o aux | Remove auxiliary centers; $K^{\mathrm{uni}}$ reduces to the main-center rule set. |
| w/o aux transfer | Keep auxiliary centers but set $\lambda_1^{(a)} = \lambda_3^{(a)} = 0$. |
| w/o $K^{\mathrm{src}}$ | Skip source-pool screening by setting $K^{\mathrm{src}} = K^{\mathrm{uni}}$. |
| w/o $K^t$ | Skip target-side rule identification by setting $K^t = K^{\mathrm{src}}$. |
| target-only $K^t$ | Use target support only in $K^t$ selection by setting $\beta_t = 1$. |
| w/o $p$-fusion | Remove the rule-level consequent prior by setting $\lambda_1^t = 0$. |
| w/o $r$-fusion | Remove the response-level anchor by setting $\lambda_3^t = 0$. |

Table S11 reports the complete dataset-wise accuracies for Full MCRT-TSK ($q = 2$) and all variants, and Table S12 gives the exact full-minus-variant deltas underlying main-text Fig. 2(C–D). Fig. 2(D) visualizes the same delta matrix. In accuracy-fraction units, the mean losses are 0.0657 without auxiliary information, 0.0579 without auxiliary transfer, 0.0408 without $K^{\mathrm{src}}$, 0.0385 without $K^t$, 0.0529 with target-only $K^t$, 0.0570 without $p$-fusion, and 0.0572 without $r$-fusion. These values support the aggregate importance of the auxiliary transfer path and the two fusion terms; zero or negative dataset-level entries reflect heterogeneous component effects rather than a universal improvement claim.

Table S11: Dataset-wise ablation accuracy (ACC↑, %; mean (standard deviation)); boldface and underlining indicate the row-wise best and second-best variants, respectively. Means are shown above their parenthesized standard deviations.

| | (a) Auxiliary information and source screening | | | | (b) Target screening and fusion priors | | | |
|---|---|---|---|---|---|---|---|---|
| Abbr. | Full MCRT-TSK | w/o aux | w/o aux transfer | w/o $K^{\mathrm{src}}$ | w/o $K^{t}$ | target-only $K^{t}$ | w/o $p$-fusion | w/o $r$-fusion |
| Son | **85.00 (11.49)** | 68.75 (16.93) | 71.25 (19.59) | 70.00 (14.67) | 68.75 (19.76) | 72.50 (20.24) | 67.50 (18.82) | 70.00 (18.82) |
| Hea | **88.33 (7.03)** | 83.33 (6.80) | 79.17 (8.10) | 80.00 (8.96) | 78.33 (9.78) | 83.33 (5.56) | 82.50 (9.17) | 80.00 (7.03) |
| Bre | **96.82 (3.07)** | 94.55 (4.18) | 94.55 (4.18) | 94.55 (3.59) | 95.00 (3.98) | 94.55 (4.18) | 95.00 (3.35) | 94.09 (3.74) |
| Aus | **90.74 (3.60)** | 78.15 (14.65) | 76.30 (15.04) | 78.15 (12.76) | 79.63 (13.66) | 75.19 (14.72) | 77.41 (13.46) | 74.81 (14.79) |
| Blo | **82.33 (4.17)** | 76.67 (1.57) | 76.67 (0.00) | 77.00 (1.05) | 77.00 (1.05) | 77.00 (1.05) | 76.33 (1.05) | 77.00 (1.05) |
| Dia | **74.52 (4.15)** | 69.68 (6.49) | 70.97 (7.29) | 73.23 (4.57) | 72.26 (7.48) | 70.32 (5.44) | 70.32 (6.59) | 70.00 (5.49) |
| Qsa | **82.86 (2.19)** | 78.10 (7.84) | 78.33 (6.29) | 79.29 (6.64) | 79.52 (6.85) | 78.81 (7.57) | 78.57 (7.01) | 78.57 (7.53) |
| DiR | **65.22 (2.29)** | 61.30 (6.86) | 58.70 (7.74) | 60.43 (6.38) | 60.43 (7.16) | 60.87 (5.32) | 57.17 (8.07) | 60.65 (6.10) |
| Ban | 95.93 (3.24) | 83.70 (11.44) | 88.52 (9.19) | 96.67 (3.24) | **96.85 (3.71)** | 87.59 (9.99) | 88.89 (8.95) | 88.52 (9.51) |
| Spa | **87.14 (1.08)** | 84.86 (3.45) | 83.95 (3.49) | 84.32 (3.32) | 84.32 (3.32) | 84.70 (3.41) | 84.43 (3.32) | 84.81 (3.45) |
| Iri | **100.00 (0.00)** | 80.00 (18.92) | 86.67 (15.32) | 96.67 (10.54) | 95.00 (11.25) | 86.67 (17.21) | 90.00 (16.10) | 85.00 (18.34) |
| Win | **100.00 (0.00)** | 91.43 (12.05) | 92.86 (10.10) | 95.71 (9.64) | **100.00 (0.00)** | 97.14 (6.02) | 94.29 (9.99) | 95.71 (9.64) |
| Con | **53.28 (1.71)** | 47.59 (5.22) | 48.45 (4.77) | 47.76 (4.46) | 50.52 (5.63) | 47.59 (5.22) | 48.45 (6.21) | 48.28 (4.06) |
| Car | **78.84 (3.36)** | 75.94 (4.64) | 75.65 (3.97) | 77.10 (3.97) | 77.68 (3.07) | 76.52 (5.06) | 75.94 (5.12) | 77.39 (5.08) |
| WiQ | **59.50 (2.08)** | 56.91 (4.40) | 58.45 (3.25) | 58.73 (2.72) | 58.45 (3.07) | 57.62 (3.42) | 57.07 (3.17) | 57.85 (3.61) |
| WiD | **55.08 (2.14)** | 52.15 (3.20) | 54.56 (3.73) | 53.79 (3.13) | 53.44 (2.70) | 53.23 (3.13) | 52.77 (2.99) | 53.33 (3.31) |
| Wav | **84.55 (1.01)** | 79.80 (3.21) | 81.95 (3.50) | 84.30 (2.46) | **84.55 (2.28)** | 82.40 (3.54) | 81.80 (3.75) | 82.20 (3.44) |
| Pag | **94.00 (0.47)** | 92.95 (0.92) | 92.86 (1.15) | 92.95 (1.01) | 93.00 (1.10) | 92.86 (0.98) | 93.05 (1.05) | 92.86 (1.07) |

Table S12: Dataset-wise accuracy deltas for main-text Fig. 2(D), computed as Full MCRT-TSK ($q = 2$) minus each ablated variant in accuracy-fraction units (0.01 equals one percentage point). Rows are ordered by task type and then by increasing sample size $N$. Positive values indicate that the ablated variant is worse; the final row gives the displayed-dataset mean.

| | (a) Auxiliary information and screening | | | | (b) Target-only scoring and fusion priors | | |
|---|---|---|---|---|---|---|---|
| Abbr. | w/o aux | w/o aux transfer | w/o $K^{\mathrm{src}}$ | w/o $K^t$ | target-only $K^t$ | w/o $p$-fusion | w/o $r$-fusion |
| Son | 0.1625 | 0.1375 | 0.1500 | 0.1625 | 0.1250 | 0.1750 | 0.1500 |
| Hea | 0.0500 | 0.0917 | 0.0833 | 0.1000 | 0.0500 | 0.0583 | 0.0833 |
| Bre | 0.0227 | 0.0227 | 0.0227 | 0.0182 | 0.0227 | 0.0182 | 0.0273 |
| Aus | 0.1259 | 0.1444 | 0.1259 | 0.1111 | 0.1556 | 0.1333 | 0.1593 |
| Blo | 0.0567 | 0.0567 | 0.0533 | 0.0533 | 0.0533 | 0.0600 | 0.0533 |
| Dia | 0.0484 | 0.0355 | 0.0129 | 0.0226 | 0.0419 | 0.0419 | 0.0452 |
| Qsa | 0.0476 | 0.0452 | 0.0357 | 0.0333 | 0.0405 | 0.0429 | 0.0429 |
| DiR | 0.0391 | 0.0652 | 0.0478 | 0.0478 | 0.0435 | 0.0804 | 0.0457 |
| Ban | 0.1222 | 0.0741 | -0.0074 | -0.0093 | 0.0833 | 0.0704 | 0.0741 |
| Spa | 0.0227 | 0.0319 | 0.0281 | 0.0281 | 0.0243 | 0.0270 | 0.0232 |
| Iri | 0.2000 | 0.1333 | 0.0333 | 0.0500 | 0.1333 | 0.1000 | 0.1500 |
| Win | 0.0857 | 0.0714 | 0.0429 | 0.0000 | 0.0286 | 0.0571 | 0.0429 |
| Con | 0.0569 | 0.0483 | 0.0552 | 0.0276 | 0.0569 | 0.0483 | 0.0500 |
| Car | 0.0290 | 0.0319 | 0.0174 | 0.0116 | 0.0232 | 0.0290 | 0.0145 |
| WiQ | 0.0260 | 0.0105 | 0.0077 | 0.0105 | 0.0188 | 0.0243 | 0.0166 |
| WiD | 0.0292 | 0.0051 | 0.0128 | 0.0164 | 0.0185 | 0.0231 | 0.0174 |
| Wav | 0.0475 | 0.0260 | 0.0025 | 0.0000 | 0.0215 | 0.0275 | 0.0235 |
| Pag | 0.0105 | 0.0114 | 0.0105 | 0.0100 | 0.0114 | 0.0095 | 0.0114 |
| Mean | 0.0657 | 0.0579 | 0.0408 | 0.0385 | 0.0529 | 0.0570 | 0.0572 |

# D Heterogeneity, Target-Budget, and Natural-Center Stress Tests of the Shared-Antecedent Interface

## D.1 Evaluation Settings

These stress tests examine two complementary mechanism questions: whether shared antecedent semantics remain useful under center heterogeneity, and whether transferred consequent blocks benefit from preserving their rule identities. The evaluation comprises three settings: controlled heterogeneity, target-budget variation, and natural Heart-site partitions. Controlled heterogeneity is evaluated on six datasets under one IID baseline and six shifted conditions: mild, moderate, and severe feature heterogeneity and mild, moderate, and severe label heterogeneity, with $\alpha \in \{\infty, 10, 1, 0.3\}$ denoting IID, mild, moderate, and severe allocation, respectively. Target-budget results are obtained under moderate feature heterogeneity with $k \in \{5, 10, 20\}$ labeled target samples per class. Natural-center results are reported for four UCI Heart sites with $k = 5$. Table S13 summarizes these settings.

For controlled feature heterogeneity, samples are divided into six feature-space groups, and each group is assigned to four centers according to $\boldsymbol{p}_g \sim \mathrm{Dirichlet}(\alpha \boldsymbol{r})$. Controlled label heterogeneity uses class-specific allocations $\boldsymbol{p}_c \sim \mathrm{Dirichlet}(\alpha \boldsymbol{r})$, where $\boldsymbol{r} = (0.30, 0.25, 0.25, 0.20)$. The target training set contains exactly $k$ labeled samples per class.

The four processed UCI Heart datasets are treated as separate sites with 13 aligned features and a binary diagnosis label defined as 0 versus $> 0$. Each site serves once as the target center. Five samples per class form the target training set, and the remaining target samples form the target test set. The largest non-target site serves as the main center, and the other two sites serve as auxiliary centers. Shared configurations use median imputation and min–max scaling estimated from the pooled training data, whereas Local-Target uses statistics estimated only from the target training set.

Table S13: Evaluation settings for controlled heterogeneity, target-budget variation, and natural Heart-site partitions.

| Task family | Datasets or sites | Conditions | Target labels per class | Seeds | Independent block |
|---|---|---|---|---|---|
| Controlled heterogeneity | 6 datasets | IID; mild, moderate, and severe feature and label heterogeneity | 5 | 2002–2011 | Dataset |
| Target budget | 5 datasets | Moderate feature shift | 5, 10, 20 | 2002–2011 | Dataset |
| Natural Heart sites | 4 sites | Each site used once as target | 5 | 3002–3011 | Target site |

### D.2 Mechanism-Isolation Comparison and Statistical Measures

The stress-test evaluation includes six diagnostic configurations and two mechanism-directed contrasts. Local-Target versus Shared-Antecedent isolates the structural value of shared antecedent semantics. Full versus Shuffled-Rule-Prior tests whether the rule-level consequent prior must preserve antecedent identity. Panel (e) of main-text Fig. 3(A) visualizes the nine mean Full-minus-Shuffled margins, while Table S15 reports the corresponding values, positive-block counts, and exact tests.

**Local-Target** estimates both target antecedents and target consequents using only the labeled target training data.

**Shared-Antecedent** uses the main-center antecedent partition and the resulting shared identity space, selects target-supported rules from this space, and estimates all target consequents using only the labeled target training data. It does not construct $K^{\mathrm{src}}$, $\mathbf{P}^t_{\mathrm{fus}}$, or $\mathbf{r}^t_{\mathrm{fus}}$.

**Shuffled-Rule-Prior** starts from the Full configuration and preserves $K^t$, the target training and test data, the number and norm of rows in $\mathbf{P}^t_{\mathrm{fus}}$, $\mathbf{r}^t_{\mathrm{fus}}$, and all target regularizers. It permutes the rows of $\mathbf{P}^t_{\mathrm{fus}}$ within each rule order, thereby breaking only the rule-identity–consequent correspondence.

The first pairwise comparison tests whether shared antecedent semantics provide a useful cross-center structural interface without source-consequent regularization. The second is an identity-breaking negative control for the rule-level consequent prior. Neither contrast is interpreted as an unconditional claim about every component of the complete objective.

Results are reported using ACC, Macro-F1, and G-mean. Datasets are treated as independent blocks in the controlled-heterogeneity and target-budget comparisons, and target sites are treated as independent blocks in the Heart comparison. Exact sign-flip tests are applied to block-level paired means, with Benjamini–Hochberg correction within each prespecified metric family.

### D.3 Controlled-Heterogeneity Diagnostics

Panels (a,b) of main-text Fig. 3(A) and Table S14 report the differences between the severe and IID conditions. Feature $\mathrm{MMD}^2$, feature ASMD, domain-classifier balanced accuracy, and label TVD are higher under the severe condition for all six dataset blocks. Each diagnostic has 6/6 positive blocks, with an exact two-sided $p$-value and a BH-adjusted $q$-value of 0.03125. These results verify that the stress generator produces the intended feature- and label-distribution shifts.

### D.4 Shared-Antecedent Results

Relative to Local-Target, Shared-Antecedent yields positive mean gains in all nine task-family–metric combinations. Under controlled heterogeneity, the mean gains in ACC, Macro-F1, and G-mean are $+0.9$, $+1.8$, and $+1.2$ percentage points (pp). Across target budgets, the corresponding gains are $+0.1$, $+0.2$, and $+1.0$ pp. Across the four natural UCI Heart sites, the gains are $+2.6$, $+1.8$, and $+1.5$ pp. Among the six stress-test configurations, Shared-Antecedent ranks first on 75% of the Heart sites for ACC and on 100% for both Macro-F1 and G-mean. These gains support the

Table S14: Severe-minus-IID differences for the four controlled-heterogeneity diagnostics. Confidence intervals are calculated over dataset blocks.

| Diagnostic | Comparison | Mean diff. | 95% CI | Positive blocks | Exact $p$ | BH $q$ |
|---|---|---|---|---|---|---|
| Feature $MMD^2$ | Feature severe − IID | 0.1621 | [0.1024, 0.2238] | 6/6 | 0.03125 | 0.03125 |
| Feature ASMD | Feature severe − IID | 0.4300 | [0.2977, 0.5816] | 6/6 | 0.03125 | 0.03125 |
| Domain BACC | Feature severe − IID | 0.1709 | [0.1477, 0.1936] | 6/6 | 0.03125 | 0.03125 |
| Label TVD | Label severe − IID | 0.4153 | [0.2468, 0.5694] | 6/6 | 0.03125 | 0.03125 |

**(a)** Controlled dataset gains (Shared - Local, pp)

| | Acc. | F1 | GM |
|---|---|---|---|
| Waveform | +13.3 | +13.8 | +15.7 |
| Car | +0.6 | -0.3 | -2.7 |
| Diabetes | +9.6 | +8.6 | +8.7 |
| Page | -19.9 | -11.4 | -11.5 |
| Sonar | +6.6 | +7.2 | +11.0 |
| Spambase | -5.0 | -7.4 | -13.9 |

**(b)** Target-budget subset gains (post hoc, pp)

| | Acc. | F1 | GM |
|---|---|---|---|
| Waveform, k=5 | +10.9 | +14.2 | +21.2 |
| Waveform, k=10 | +2.2 | +4.9 | +8.9 |
| Waveform, k=20 | +0.7 | +1.4 | +5.4 |
| Diabetes, k=5 | +2.9 | +4.6 | +5.7 |
| Diabetes, k=10 | +3.4 | +2.2 | +0.8 |
| Diabetes, k=20 | +0.8 | +1.0 | +2.9 |

Figure S1: Dataset-level Shared-Antecedent minus Local-Target differences in percentage points: (a) results averaged over the seven controlled conditions; (b) post-hoc target-budget subset results for Waveform and Diabetes. The comparison isolates shared antecedent semantics and excludes source-consequent priors.

robustness of the shared-antecedent semantic interface, but they should not be attributed to the rule-level fusion prior or the response-level anchor, both of which are excluded from this configuration.

Figure S1(a) reports dataset-level Shared-Antecedent minus Local-Target differences averaged over the seven controlled conditions. Waveform, Diabetes, and Sonar show positive mean gains for all three metrics, Car shows a mixed response, and Page Blocks and Spambase show negative mean differences for all three metrics. Figure S1(b) provides post-hoc target-budget detail for Waveform and Diabetes; every displayed budget-specific difference is positive across ACC, Macro-F1, and G-mean.

### D.5 Identity-Preserving Consequent-Prior Control

Table S15 reports the Full minus Shuffled-Rule-Prior contrast. All nine task-family–metric mean differences are positive. The comparison is strongest for Macro-F1 and G-mean in the controlled and target-budget families, where preserving rule identity improves the mean balanced metrics by 2.30–3.03 pp. Because the control changes only the correspondence between rule identity and consequent prior, these consistently positive margins isolate the benefit of identity-preserving consequent reuse.

### D.6 Natural Heart-Site Results

The Heart rotations exhibit measurable natural-center mismatch. Across target-site rotations, mean pairwise ASMD ranges from 0.525 to 0.665, $MMD^2$ from 0.124 to 0.165, label TVD from 0.199 to 0.328, and domain-classifier BACC from 0.699 to 0.845. Mean source-rule Jaccard is only 0.632–0.659, confirming that the site-specific source rule sets overlap only partially rather than being identical.

Table S16 reports the Shared-Antecedent minus Local-Target differences for the four Heart target sites. The mean differences across sites are $+2.62$ pp in ACC, $+1.77$ pp in Macro-F1, and $+1.48$ pp in G-mean.

Table S15: Identity-preserving consequent-prior control. Entries are Full minus Shuffled-Rule-Prior mean differences in percentage points. Controlled effects first average seeds and conditions within each dataset; target-budget effects first average seeds and budgets within each dataset; Heart effects first average seeds within each target site. Positive values favor preservation of the rule-identity–consequent correspondence.

| Task family | Metric | Mean $\Delta$ (pp) | Positive | Exact $p$ |
|---|---|---|---|---|
| Controlled heterogeneity | ACC | +2.08 | 3/6 | 0.375 |
| Controlled heterogeneity | Macro-F1 | +2.64 | 4/6 | 0.125 |
| Controlled heterogeneity | G-mean | +2.30 | 3/6 | 0.250 |
| Target budget | ACC | +1.86 | 3/5 | 0.250 |
| Target budget | Macro-F1 | +2.74 | 3/5 | 0.250 |
| Target budget | G-mean | +3.03 | 3/5 | 0.250 |
| Natural Heart sites | ACC | +1.54 | 2/4 | 0.500 |
| Natural Heart sites | Macro-F1 | +1.38 | 2/4 | 0.500 |
| Natural Heart sites | G-mean | +3.85 | 2/4 | 0.500 |

Table S16: Shared-Antecedent minus Local-Target differences on the four natural Heart target sites.

| Target site | $\Delta$ACC (pp) | $\Delta$Macro-F1 (pp) | $\Delta$G-mean (pp) |
|---|---|---|---|
| Cleveland | +0.92 | +0.88 | +0.81 |
| Hungary | +2.71 | +1.73 | +0.07 |
| Switzerland | +8.23 | +4.12 | +2.49 |
| VA Long Beach | $-1.37$ | +0.35 | +2.55 |

# E Supplementary Analysis and Discussion Details

## E.1 Rule-Complexity Analysis

Rule complexity is used as a retained-model-size diagnostic for the compared TSK-family implementations. It is not treated as a direct accuracy criterion, and it is not a hardware-independent runtime bound. The reason is that the compared methods expose different rule-base parameterizations: some methods retain class-specific fuzzy rules, some use integrated or transferred rule bases, and MIP-TRL-FS exposes a projection-matrix proxy rather than an identical rule-count object. Therefore, the values below should be interpreted as implementation-aligned retained-size evidence, with missing cells preserved as unavailable evidence rather than converted to zeros. In this context, main-text Fig. 3(D) and the supplementary complexity tables are diagnostic summaries, not a strict unified ranking of computational complexity across heterogeneous model parameterizations.

Let $\mathcal{C}_{m,D,r}$ be the complexity recorded for method $m$, dataset $D$, and valid repeat $r$. The displayed dataset-wise value is

$$\bar{\mathcal{C}}_{m,D} = \frac{1}{|\mathcal{V}_{m,D}|} \sum\nolimits_{r \in \mathcal{V}_{m,D}} \mathcal{C}_{m,D,r},$$

where $\mathcal{V}_{m,D}$ contains only valid repeats. For MCRT-TSK, the implementation records the target retained-rule set $K^t$. If $N_c$ is the number of classes and $o_\kappa$ is the antecedent order of rule $\kappa$, the recorded complexity is

$$\mathcal{C}_{\mathrm{MCRT}} = \sum\nolimits_{\kappa \in K^t} N_c(o_\kappa + 1),$$

where the $+1$ term counts one response parameter for each class. Under the default $q = 2$, $o_\kappa \in \{1, 2\}$. This definition matches the retained-rule complexity computed by the MCRT-TSK target and auxiliary training routines, and it explains why high-dimensional datasets can produce large upper ranges: second-order rule construction expands the retained rule-identity space before target-side screening.

For conventional full TSK rule bases used by several baselines, the diagnostic follows the implementation-exposed rule base parameter count. With an antecedent-center matrix $G$, conse-

quent matrix $P_g$, and $N_c$ classes, the common recorded form is

$$\mathcal{C}_{\mathrm{TSK}} = N_c|G| + |P_g|.$$

For class-specific or integrated rule bases, the corresponding class-specific counts are summed across classes. For MIP-TRL-FS, the released aligned runner exposes the proxy $\mathcal{C}_{\mathrm{MIP}} = \sum_v |P_v|$, where $v$ indexes learned views. These definitions make the comparison useful as a diagnostic of retained representation size, but they also require caution: a lower value does not by itself imply a better transfer model, and the complexity table should be read jointly with ACC and valid-run coverage.

Table S17: Method-level rule-complexity and training-time summaries over valid displayed datasets. Training time is reported in seconds. Valid gives the common number of valid displayed datasets for both diagnostics.

| | | Rule complexity | | | Training time (s) | | |
|---|---|---|---|---|---|---|---|
| Method | Valid | Mean | Median | Max | Mean | Median | Max |
| 0-TSK | 18 | 887.7 | 725.0 | 2226.0 | 0.0242 | 0.0111 | 0.0929 |
| 1-TSK | 18 | 609.9 | 460.5 | 1936.0 | 0.0101 | 0.0061 | 0.0334 |
| TL-SSL-TSK | 9 | 1046.8 | 945.0 | 2196.0 | 0.1295 | 0.1094 | 0.3480 |
| ETTL-TSK-FS | 10 | 1558.0 | 1076.0 | 5060.0 | 2.2964 | 0.3382 | 19.777 |
| FL-TSK-FRBSs | 10 | 1277.3 | 745.6 | 3794.2 | 0.1331 | 0.1269 | 0.1502 |
| MKTC-R0T | 10 | 1609.7 | 1197.0 | 5133.2 | 0.0001 | 0.0001 | 0.0004 |
| MIP-TRL-FS | 18 | 2131.2 | 1167.6 | 10126.0 | 0.7636 | 0.0886 | 3.4651 |
| MTSK-DA | 13 | 511.9 | 356.4 | 1159.0 | 10275.98 | 5793.40 | 34904.06 |
| KCF-LSG-MTSK | 16 | 696.5 | 481.0 | 1694.0 | 309.58 | 81.510 | 1512.41 |
| MCRT-TSK (ours) | 18 | 4827.1 | 1361.4 | 29223.4 | 1.6248 | 0.2022 | 13.061 |

Table S18 gives the dataset-wise retained-size matrix without best-value marking. MCRT-TSK can retain a larger rule-identity space because it constructs a bounded second-order shared candidate space before source screening and target filtering. These values should be read as implementation-aligned structural diagnostics together with runtime and valid-result coverage, not as a unified complexity ranking.

Table S18: Dataset-wise rule complexity (↓; means over valid repeats). Dashes denote unsupported, failed, or timed-out evidence.

| Data | 0-TSK | 1-TSK | TL-SSL -TSK | ETTL-TSK -FS | FL-TSK -FRBSs | MKTC -R0T | MIP-TRL -FS | MTSK -DA | KCF-LSG -MTSK | MCRT-TSK (ours) |
|---|---|---|---|---|---|---|---|---|---|---|
| Son | 1464.0 | 1936.0 | – | 3146.0 | 3025.0 | 2580.2 | 10126.0 | 1064.8 | 1694.0 | 29223.4 |
| Hea | 238.0 | 432.0 | 945.0 | 837.0 | 1009.8 | 473.6 | 1484.0 | 291.6 | 216.0 | 716.8 |
| Bre | 1147.0 | 915.0 | 2196.0 | 2196.0 | 3794.2 | 1909.4 | 4154.0 | 1159.0 | 1146.8 | 6141.6 |
| Aus | 375.0 | 435.0 | 1247.0 | 1073.0 | 1293.4 | 912.0 | 1260.0 | 348.0 | 261.0 | 954.4 |
| Blo | 170.0 | 396.0 | 351.0 | 351.0 | 153.0 | 396.0 | 224.0 | 399.6 | 363.6 | 92.4 |
| Dia | 360.0 | 204.0 | 544.0 | 323.0 | 156.4 | 528.0 | 662.4 | 319.6 | 265.2 | 340.0 |
| Qsa | 1890.0 | 913.0 | 1743.0 | 1079.0 | 481.4 | 2080.8 | 4536.0 | 1012.6 | 1195.2 | 10618.0 |
| DiR | 820.0 | 546.0 | 858.0 | 1092.0 | 2316.6 | 1482.0 | 2068.0 | 631.8 | 397.8 | 2251.2 |
| Ban | 230.0 | 108.0 | 387.0 | 423.0 | 313.2 | 602.0 | 128.0 | 325.8 | 298.8 | 108.6 |
| Spa | 2146.0 | 1150.0 | 1150.0 | 5060.0 | 230.0 | 5133.2 | 6612.0 | – | – | 22790.8 |
| Iri | 75.0 | 135.0 | – | – | – | – | 144.0 | 135.0 | 113.4 | 158.1 |
| Win | 231.0 | 486.0 | – | – | – | – | 1593.2 | 356.4 | 307.8 | 1874.7 |
| Con | 540.0 | 285.0 | – | – | – | – | 864.0 | 444.6 | 564.3 | 619.5 |
| Car | 630.0 | 416.0 | – | – | – | – | 470.4 | 166.4 | 1279.2 | 494.0 |
| WiQ | 846.0 | 379.5 | – | – | – | – | 804.0 | – | 1635.3 | 1257.3 |
| WiD | 2226.0 | 966.0 | – | – | – | – | 1075.2 | – | 644.0 | 2839.2 |
| Wav | 1188.0 | 645.0 | – | – | – | – | 1408.0 | – | 761.1 | 4942.5 |
| Pag | 1402.5 | 630.0 | – | – | – | – | 748.0 | – | – | 1465.5 |

### E.2 TRAINING-TIME ANALYSIS

Training time is reported as an empirical wall-clock diagnostic for the aligned MATLAB implementations. It is not a hardware-independent theoretical bound, because it depends on implementation choices, search spaces, stopping behavior, data preprocessing, and the status coverage of each method. The target-only diagnostic baselines represent single-center reference costs, whereas MCRT-TSK and the multi-source baselines include their corresponding rule construction, transfer, calibration, graph, or alignment stages under the main-text protocol.

Let $T_{m,D,r}$ denote the elapsed training time, in seconds, recorded for method $m$ on dataset $D$ and valid repeat $r$. The displayed dataset-wise value is

$$\bar{T}_{m,D} = \frac{1}{|\mathcal{V}_{m,D}|} \sum\nolimits_{r \in \mathcal{V}_{m,D}} T_{m,D,r},$$

where $\mathcal{V}_{m,D}$ contains only valid repeats. Unsupported multi-class cases, failed runs, and timeout entries are excluded rather than imputed. This rule prevents missing evidence from being silently converted into a numerical advantage or penalty.

The training-time columns of Table S17 and the dataset-wise values in Table S19 complement Fig. S2 by reporting the exact method-level and dataset-wise training times. In the reported aligned MATLAB runs, MCRT-TSK is slower than the target-only diagnostic baselines and MKTC-R0T, but faster than the valid KCF-LSG-MTSK and MTSK-DA runs summarized here. MTSK-DA has a median of 5793.40 seconds over 13 valid datasets, and KCF-LSG-MTSK has a median of 81.510 seconds over 16 valid datasets, whereas MCRT-TSK has a median of 0.2022 seconds over all 18 displayed datasets. The detailed table below preserves this coverage distinction explicitly.

Table S19: Dataset-wise training time in seconds (↓; means over valid repeats). Dashes denote unsupported, failed, or timed-out evidence.

| Data | 0-TSK | 1-TSK | TL-SSL -TSK | ETTL-TSK -FS | FL-TSK -FRBSs | MKTC -R0T | MIP-TRL -FS | MTSK -DA | KCF-LSG -MTSK | MCRT-TSK (ours) |
|---|---|---|---|---|---|---|---|---|---|---|
| Son | 0.0031 | 0.0077 | – | 0.0996 | 0.1489 | 0.0001 | 0.0938 | 791.88 | 132.97 | 9.1255 |
| Hea | 0.0018 | 0.0024 | 0.0699 | 0.0580 | 0.1481 | 0.0001 | 0.0104 | 776.96 | 10.757 | 0.1041 |
| Bre | 0.0095 | 0.0067 | 0.2383 | 0.2041 | 0.1214 | 0.0001 | 0.0497 | 5674.90 | 90.316 | 1.0581 |
| Aus | 0.0033 | 0.0035 | 0.1237 | 0.4407 | 0.1219 | 0.0004 | 0.0292 | 5793.40 | 22.400 | 0.3582 |
| Blo | 0.0044 | 0.0053 | 0.0254 | 0.1977 | 0.1184 | 0.0001 | 0.0248 | 7647.75 | 33.416 | 0.0331 |
| Dia | 0.0072 | 0.0044 | 0.0366 | 0.2356 | 0.1208 | 0.0001 | 0.0299 | 4310.14 | 27.832 | 0.0480 |
| Qsa | 0.0252 | 0.0089 | 0.3480 | 0.4912 | 0.1482 | 0.0001 | 0.1386 | 14049.67 | 163.32 | 3.5755 |
| DiR | 0.0148 | 0.0055 | 0.1094 | 0.6557 | 0.1318 | 0.0001 | 0.0834 | 10398.22 | 57.300 | 0.3553 |
| Ban | 0.0075 | 0.0051 | 0.0439 | 0.8041 | 0.1208 | 0.0001 | 0.0831 | 19918.32 | 72.703 | 0.0334 |
| Spa | 0.0929 | 0.0334 | 0.1705 | 19.777 | 0.1502 | 0.0001 | 2.0734 | – | – | 13.061 |
| Iri | 0.0011 | 0.0011 | – | – | – | – | 0.0026 | 216.95 | 5.1229 | 0.0144 |
| Win | 0.0015 | 0.0018 | – | – | – | – | 0.0086 | 596.40 | 10.111 | 0.2668 |
| Con | 0.0127 | 0.0056 | – | – | – | – | 0.1406 | 28509.07 | 184.96 | 0.0961 |
| Car | 0.0210 | 0.0147 | – | – | – | – | 0.2360 | 34904.06 | 178.85 | 0.0326 |
| WiQ | 0.0579 | 0.0146 | – | – | – | – | 1.7041 | – | 1512.41 | 0.2065 |
| WiD | 0.0640 | 0.0207 | – | – | – | – | 3.3823 | – | 953.50 | 0.1979 |
| Wav | 0.0580 | 0.0266 | – | – | – | – | 2.1890 | – | 1497.28 | 0.5736 |
| Pag | 0.0492 | 0.0139 | – | – | – | – | 3.4651 | – | – | 0.1057 |

Figure S2 provides a task-aware visual counterpart to the training-time tables. Panel (a) reports average ACC ranks separately for binary and multi-class evidence, avoiding a pooled rank over methods with different valid-dataset coverage. Panel (b) places mean ACC against log-scale mean training time within each task type, so the upper-left region indicates a more favorable accuracy–cost trade-off. The figure follows the same valid-entry rule as the tables: unsupported, failed, and timed-out runs are not imputed.

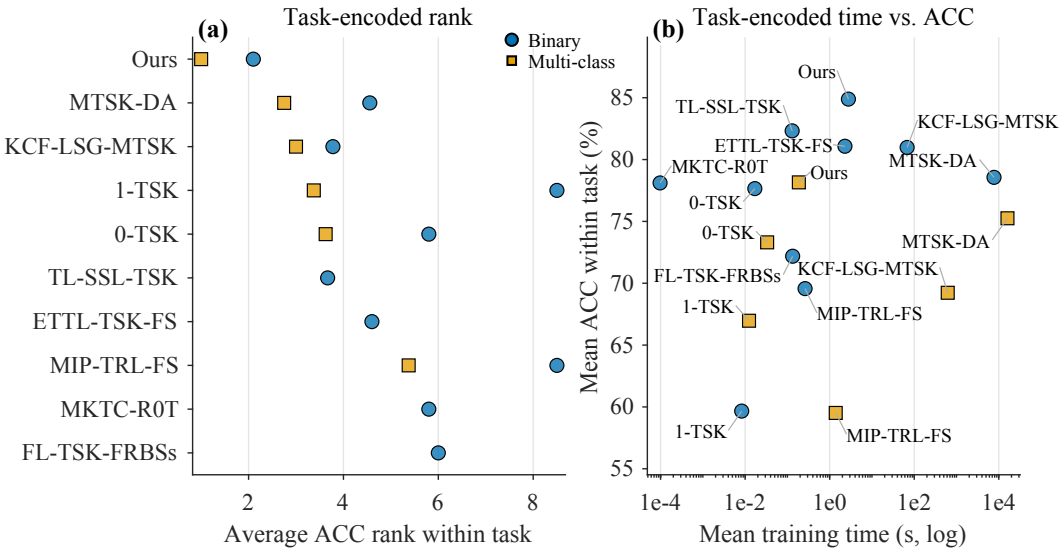


Figure S2: Supplementary task-encoded rank and time diagnostic. Panel (a) overlays protocol-matched average accuracy ranks for binary and multi-class evidence on a shared method axis. Panel (b) overlays task-specific mean ACC (%) and log-scale mean training time. Blue circles denote binary datasets and orange squares denote multi-class datasets; all ranks and means are computed within each task type.

### E.3 Rule-Level Interpretability and Linguistic Rendering Details

This section expands the rule-level evidence behind the main-text audit. The Diabetes target model illustrates the complete trace, while Table S20 reports feature-name coverage, reconstructed target-rule cardinality, local decision margin, and the leading positive rule contribution for additional binary datasets. These diagnostics describe model traceability rather than predictive performance.

At the retained-rule level, the audit follows the model objects defined in the method section: MCRT-TSK fixes shared antecedent semantics, screens $K^{\mathrm{uni}}$ into $K^{\mathrm{src}}$, selects $K^t$ with $\hat{r}^t = \beta_t s^t + (1 - \beta_t) r^{\mathrm{src}}$, aligns matched source consequents, and solves the final target consequent matrix. Figure S3 presents the full six-panel visual trace summarized in main-text Fig. 3(C).

The linguistic labels in the antecedent panels are generated from the ordered fuzzy centers of each feature partition: smaller normalized centers receive lower-level labels such as “very low” or “low”, whereas larger centers receive higher-level labels. Thus, documented UCI feature names provide the domain semantics, while the level labels are automatic partition-order descriptors rather than hand-written expert rules. For encoded or categorical attributes, such labels should be read only as levels of the numeric encoding, not as clinical or physical thresholds. This follows the common fuzzy-system practice of rendering a learned rule as an IF clause over linguistic fuzzy sets and a THEN clause over model responses. For a zero-order TSK classifier, however, the THEN part is a vector of class responses; we therefore verbalize it only as a class-response tendency, not as a calibrated probability or causal statement.

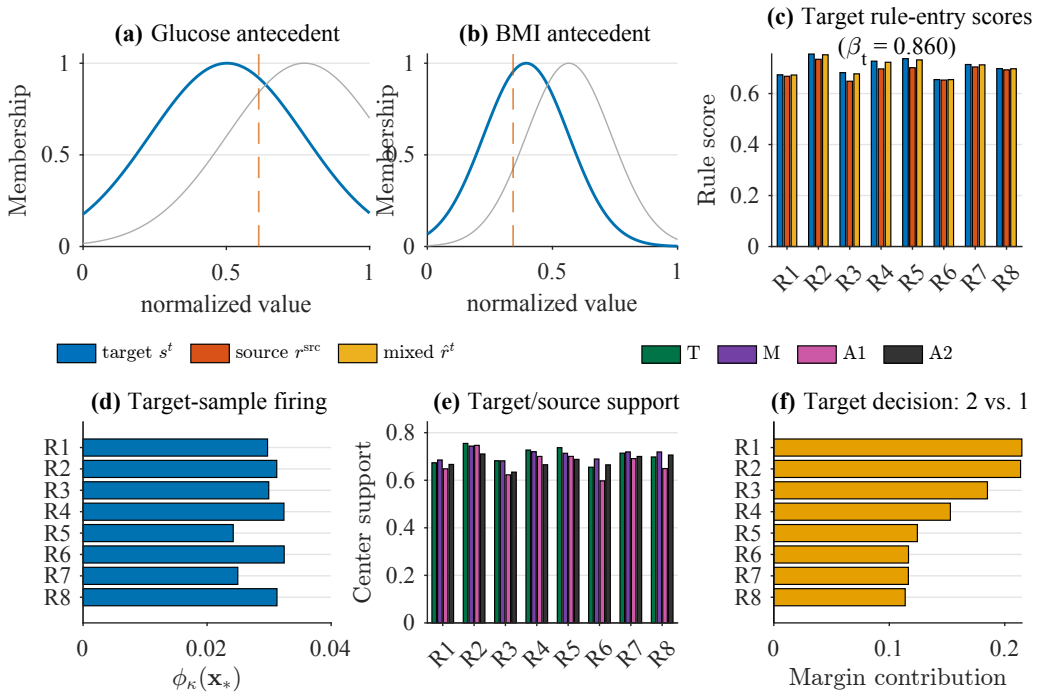


Figure S3: Detailed rule-level traceability audit for the reconstructed Diabetes target test sample. Panels (a) and (b) show named-feature fuzzy antecedent semantics for the leading margin-contributing rules. Panel (c) reports target-center support, source-side support proxy, and mixed target relevance used to retain rules in $K^t$. Panel (d) reports target-sample firing strengths. Panel (e) contrasts target-center support with matched source-center supports. Panel (f) decomposes the predicted-vs-competing class margin into rule-level contributions.

Tables S23–S25 unpack the compact trace in main-text Fig. 3(C): antecedent semantics, target-sample firing strength $\phi_\kappa(\mathbf{x}_*)$, target-center support $s^t$, source-side support proxy $r^{\text{src}}$, mixed relevance $\hat{r}^t$, rule-identity-aligned source fusion, response-level anchor, final target consequent, and raw class-score decomposition. Table S22 adds the human-readable rendering of the highest-firing rules with positive contributions by translating the same antecedent semantics and target consequents into compact IF–THEN statements. The margin contribution is $\Delta_\kappa = \phi_\kappa(\mathbf{x}_*)\{p^t_{\kappa,\hat{y}} - p^t_{\kappa,c_{\text{opp}}}\}$. Here $s^t$ is an average support over target-center training samples, whereas $\phi_\kappa(\mathbf{x}_*)$ is the normalized firing strength for this single target test sample. In the Diabetes reconstruction, $N_t = 123$, $n_0 = 20$, and $\beta_t = N_t/(N_t + n_0) = 0.8601$, so the local rule-entry score is mainly driven by target support.

The tuned target weights are $\lambda^t_1 = 0.01$, $\lambda^t_2 = 0.01$, and $\lambda^t_3 = 0.01$. Thus, $\mathbf{p}^t_{\text{fus},\kappa}$ is a source prior/anchor in the target fit, not the final class response. The leading rule illustrates this distinction: $\mathbf{p}^t_{\text{fus},\kappa} = (-2.63, 3.56)$, whereas the final target consequent is $\mathbf{p}^t_\kappa = (-3.13, 4.09)$. In this instantiation, the displayed high-activation rules are observed in all three matched source centers; partial-overlap fusion is therefore assessed through the rule-pool screening trace rather than inferred from these local rows. Table S21 summarizes the $q = 2$ rule-pool screening trace across the 18 displayed datasets. It reports the mean sizes of the source-union rule pool $K^{\text{uni}}$, the source-supported subset $K^{\text{src}}$, and the final target-retained subset $K^t$, together with the corresponding retention ratios. These quantities document how the shared rule-identity space is progressively reduced to the target-supported rule set.

Table S20: Named-feature reconstruction diagnostics for rule-level interpretability and linguistic rendering. Each row reports a $q = 2$, repeat-1 target model with complete feature-name coverage, the reconstructed target-rule cardinality, the local decision margin, and the strongest positive rule-level margin contribution. The table contextualizes the Diabetes trace with structurally comparable binary tabular tasks; it is not an additional performance benchmark.

**(a) Model and local-decision diagnostics**

| Abbr. | $d$ | Named | $\lvert K^t \rvert$ | Margin | Top $\Delta_\kappa$ |
|---|---|---|---|---|---|
| Bre | 30 | 100% | 1136 | 0.677 | 0.021 |
| Dia | 8 | 100% | 36 | 0.576 | 0.215 |
| Blo | 4 | 100% | 19 | 0.431 | 0.158 |
| Hea | 13 | 100% | 163 | 0.095 | 0.010 |

**(b) Feature semantics**

| Dataset | Abbr. | Leading rule semantics | Feature-space note |
|---|---|---|---|
| Breast Cancer WDBC | Bre | concave points mean:medium | Diagnostic cell-nucleus descriptors with complete feature-name coverage and a dense target rule set. |
| Diabetes | Dia | pregnancies:very low & glucose:medium | Patient-level physiological attributes with complete feature-name coverage and a compact target rule set. |
| Blood Transfusion | Blo | frequency:very low | RFM-style blood-donation descriptors with complete feature-name coverage and a compact target rule set. |
| Heart Disease | Hea | exercise angina:very low & vessels:very low | Clinical attributes with ordinal/categorical numeric encodings and an intermediate target rule-set size. |

Table S21: $q = 2$ rule-pool screening trace over the 18 displayed datasets. Counts report mean rule-pool sizes, and ratios report source-supported and target-retained proportions. Source/union, target/source, and target/union denote $|K^{\mathrm{src}}|/|K^{\mathrm{uni}}|$, $|K^t|/|K^{\mathrm{src}}|$, and $|K^t|/|K^{\mathrm{uni}}|$, respectively.

| Abbr. | $\lvert K^{\mathrm{uni}} \rvert$ | $\lvert K^{\mathrm{src}} \rvert$ | $\lvert K^t \rvert$ | Source / union | Target / source | Target / union |
|---|---|---|---|---|---|---|
| Aus | 300.95 | 211.10 | 136.80 | 0.7014 | 0.6480 | 0.4546 |
| Con | 158.70 | 108.45 | 70.50 | 0.6834 | 0.6501 | 0.4442 |
| Wav | 1311.65 | 877.00 | 557.80 | 0.6686 | 0.6360 | 0.4253 |
| Ban | 40.25 | 27.35 | 16.50 | 0.6795 | 0.6033 | 0.4099 |
| Blo | 27.95 | 20.45 | 13.40 | 0.7317 | 0.6553 | 0.4794 |
| Bre | 2360.10 | 1653.10 | 1048.70 | 0.7004 | 0.6344 | 0.4443 |
| Car | 92.80 | 63.60 | 42.45 | 0.6853 | 0.6675 | 0.4574 |
| Dia | 137.85 | 96.00 | 61.45 | 0.6964 | 0.6401 | 0.4458 |
| DiR | 835.65 | 572.75 | 369.25 | 0.6854 | 0.6447 | 0.4419 |
| Hea | 282.65 | 190.75 | 120.75 | 0.6749 | 0.6330 | 0.4272 |
| Iri | 43.70 | 29.35 | 18.25 | 0.6716 | 0.6218 | 0.4176 |
| Pag | 235.25 | 159.85 | 100.25 | 0.6795 | 0.6272 | 0.4261 |
| Qsa | 3913.15 | 2743.95 | 1774.10 | 0.7012 | 0.6465 | 0.4534 |
| Son | 11230.80 | 7711.45 | 4954.50 | 0.6866 | 0.6425 | 0.4412 |
| Spa | 8558.20 | 6032.90 | 3865.50 | 0.7049 | 0.6407 | 0.4517 |
| Win | 534.65 | 352.60 | 221.60 | 0.6595 | 0.6285 | 0.4145 |
| WiD | 302.65 | 210.45 | 136.45 | 0.6954 | 0.6484 | 0.4509 |
| WiQ | 305.95 | 210.95 | 138.35 | 0.6895 | 0.6558 | 0.4522 |
| Mean | 1704.05 | 1181.78 | 758.14 | 0.6886 | 0.6402 | 0.4410 |
| Median | 301.80 | 210.70 | 136.62 | 0.6860 | 0.6416 | 0.4443 |

Table S22: Representative natural-language rendering of leading Diabetes target rules. The IF clauses are generated from documented feature semantics and ordered fuzzy-set labels; the THEN clauses summarize raw TSK target consequents as class-response tendencies, not probabilities, causal claims, or expert-authored rules.

| Rank | Field | Rendering |
|---|---|---|
| 1 | IF clause | IF number of pregnancies is very low AND glucose is medium |
| | Provenance | number of pregnancies; label from fuzzy center $c = 0.07$; sample membership $\mu = 0.95$; plasma glucose concentration; label from fuzzy center $c = 0.50$; sample membership $\mu = 0.92$ |
| | THEN clause | THEN the final target consequent favors class 2 over class 1 with raw target consequent $\mathbf{p}_\kappa^t = (-3.13, 4.09)$, response contrast 7.22; for the audited sample, this rule contributes 0.215 to the class-2-vs-class-1 margin. |
| | Local evidence | $\phi_\kappa(\mathbf{x}_*) = 0.030$, $\Delta_\kappa = 0.215$ |
| 2 | IF clause | IF glucose is medium AND diabetes pedigree function is very low |
| | Provenance | plasma glucose concentration; label from fuzzy center $c = 0.50$; sample membership $\mu = 0.92$; diabetes pedigree function; label from fuzzy center $c = 0.10$; sample membership $\mu = 0.99$ |
| | THEN clause | THEN the final target consequent favors class 2 over class 1 with raw target consequent $\mathbf{p}_\kappa^t = (-2.94, 3.90)$, response contrast 6.84; for the audited sample, this rule contributes 0.214 to the class-2-vs-class-1 margin. |
| | Local evidence | $\phi_\kappa(\mathbf{x}_*) = 0.031$, $\Delta_\kappa = 0.214$ |
| 3 | IF clause | IF glucose is medium AND BMI is low |
| | Provenance | plasma glucose concentration; label from fuzzy center $c = 0.50$; sample membership $\mu = 0.92$; body mass index; label from fuzzy center $c = 0.40$; sample membership $\mu = 0.95$ |
| | THEN clause | THEN the final target consequent favors class 2 over class 1 with raw target consequent $\mathbf{p}_\kappa^t = (-2.61, 3.56)$, response contrast 6.18; for the audited sample, this rule contributes 0.185 to the class-2-vs-class-1 margin. |
| | Local evidence | $\phi_\kappa(\mathbf{x}_*) = 0.030$, $\Delta_\kappa = 0.185$ |

Table S23: Rule-level worked example for Diabetes, repeat 1. The audited target test sample has true class 2 and predicted class 2. Rows are target rules with the largest positive contributions to the predicted-vs-competing class margin, where $\Delta_\kappa = \phi_\kappa(\mathbf{x}_*)(p_{\kappa,\hat{y}}^t - p_{\kappa,c_{\mathrm{opp}}}^t)$.

**(a) Antecedent semantics**

| Rank | Antecedent semantics |
|---|---|
| 1 | number of pregnancies is very low (F1; number of pregnancies; c=0.07, sigma=0.31, mu=0.95); glucose is medium (F2; plasma glucose concentration; c=0.50, sigma=0.27, mu=0.92) |
| 2 | glucose is medium (F2; plasma glucose concentration; c=0.50, sigma=0.27, mu=0.92); diabetes pedigree function is very low (F7; diabetes pedigree function; c=0.10, sigma=0.26, mu=0.99) |
| 3 | glucose is medium (F2; plasma glucose concentration; c=0.50, sigma=0.27, mu=0.92); BMI is low (F6; body mass index; c=0.40, sigma=0.17, mu=0.95) |
| 4 | BMI is low (F6; body mass index; c=0.40, sigma=0.17, mu=0.95); diabetes pedigree is very low (F7; diabetes pedigree function; c=0.10, sigma=0.26, mu=0.99) |
| 5 | glucose is medium (F2; plasma glucose concentration; c=0.50, sigma=0.27, mu=0.92); age is very low (F8; age; c=0.08, sigma=0.33, mu=0.77) |
| 6 | insulin is very low (F5; 2-hour serum insulin; c=0.02, sigma=0.17, mu=1.00); BMI is low (F6; body mass index; c=0.40, sigma=0.17, mu=0.95) |

**(b) Target support, decision evidence, source weights, and final consequents**

| Rank | $\phi_\kappa(\mathbf{x}_*)$ | $s^t$ | $r^{\mathrm{src}}$ | $\hat{r}^t$ | $\Delta_\kappa$ | $(\omega_M, \omega_{A1}, \omega_{A2})$ | $\mathbf{p}_\kappa^t$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.030 | 0.674 | 0.668 | 0.673 | 0.215 | (0.38, 0.33, 0.28) | (-3.13, 4.09) |
| 2 | 0.031 | 0.755 | 0.734 | 0.752 | 0.214 | (0.38, 0.33, 0.28) | (-2.94, 3.90) |
| 3 | 0.030 | 0.682 | 0.648 | 0.677 | 0.185 | (0.38, 0.33, 0.28) | (-2.61, 3.56) |
| 4 | 0.032 | 0.727 | 0.697 | 0.723 | 0.153 | (0.38, 0.33, 0.28) | (-1.88, 2.84) |
| 5 | 0.024 | 0.737 | 0.701 | 0.732 | 0.124 | (0.38, 0.33, 0.28) | (-2.08, 3.06) |
| 6 | 0.032 | 0.655 | 0.653 | 0.655 | 0.117 | (0.38, 0.33, 0.28) | (-1.33, 2.27) |

Table S24: Target- and source-center support and rule-level consequent fusion for the Diabetes worked example. The target-center support $s^t$ is reported next to matched source-center supports; $\mathbf{p}^t_{\mathrm{fus},\kappa}$ is a source-derived prior, while $\mathbf{p}^t_\kappa$ is the final target consequent after target-center learning.

**(a) Antecedent identities and support**

| Rank | Rule | $s^t$ | $s_M$ | $s_{A1}$ | $s_{A2}$ |
|---|---|---|---|---|---|
| 1 | pregnancies:very low & glucose:medium | 0.674 | 0.685 | 0.648 | 0.667 |
| 2 | glucose:medium & diabetes pedigree:very low | 0.755 | 0.744 | 0.747 | 0.711 |
| 3 | glucose:medium & BMI:low | 0.682 | 0.682 | 0.623 | 0.634 |
| 4 | BMI:low & diabetes pedigree:very low | 0.727 | 0.720 | 0.700 | 0.665 |
| 5 | glucose:medium & age:very low | 0.737 | 0.714 | 0.700 | 0.688 |
| 6 | insulin:very low & BMI:low | 0.655 | 0.689 | 0.598 | 0.665 |
| 7 | pregnancies:very low & age:very low | 0.714 | 0.719 | 0.691 | 0.700 |
| 8 | glucose:medium & insulin:very low | 0.698 | 0.719 | 0.649 | 0.706 |

**(b) Source weights and consequent vectors**

| Rank | $\omega_M$ | $\omega_{A1}$ | $\omega_{A2}$ | $\mathbf{p}^t_{\mathrm{fus},\kappa}$ | $\mathbf{p}^t_\kappa$ |
|---|---|---|---|---|---|
| 1 | 0.38 | 0.33 | 0.28 | (-2.63, 3.56) | (-3.13, 4.09) |
| 2 | 0.38 | 0.33 | 0.28 | (-3.35, 4.34) | (-2.94, 3.90) |
| 3 | 0.38 | 0.33 | 0.28 | (-2.92, 3.89) | (-2.61, 3.56) |
| 4 | 0.38 | 0.33 | 0.28 | (-0.91, 1.92) | (-1.88, 2.84) |
| 5 | 0.38 | 0.33 | 0.28 | (-2.77, 3.74) | (-2.08, 3.06) |
| 6 | 0.38 | 0.33 | 0.28 | (-0.79, 1.78) | (-1.33, 2.27) |
| 7 | 0.38 | 0.33 | 0.28 | (-0.60, 1.57) | (-1.85, 2.82) |
| 8 | 0.38 | 0.33 | 0.28 | (-2.65, 3.64) | (-1.34, 2.30) |

Table S25: **(a)** Source-level response anchor for the Diabetes worked example. The discrepancy $d_u$ is computed from the source support profile over $K^t$ and the target relevance profile $\hat{\mathbf{r}}^t_{K^t}$; $\alpha_u$ fuses the source average responses into $\mathbf{r}^t_{\mathrm{fus}}$. **(b)** Raw class-score decomposition for the audited Diabetes target test sample. Scores are TSK responses, not calibrated probabilities. The top-eight-rule sum is computed over the top eight margin-contributing rules exported in the supplementary fusion table. The prediction margin is 0.5759.

**(a) Source response anchor**

| Source | $d_u$ | $\alpha_u$ | Class-1 response | Class-2 response |
|---|---|---|---|---|
| M | 0.0145 | 0.385 | -0.365 | 0.888 |
| A1 | 0.0292 | 0.332 | -0.352 | 0.866 |
| A2 | 0.0452 | 0.283 | -0.360 | 0.882 |
| $\mathbf{r}^t_{\mathrm{fus}}$ | – | – | -0.360 | 0.879 |

**(b) Target decision**

| Class | Score | Top-8 rule sum | Status |
|---|---|---|---|
| 1 | 0.2105 | -0.5056 | |
| 2 | 0.7864 | 0.7321 | true/pred. |

### E.4 Post-hoc Audit of Empirical Cross-Center Differences

The four centers in the current benchmark are simulated by stratified random partitioning of each individual UCI dataset rather than by independent institutional acquisition. No covariate shift, label shift, concept shift, or center-specific sampling mechanism is intentionally imposed. To quantify the empirical differences that nevertheless arise from finite-sample allocation, we conduct a post-hoc descriptive audit on the center-wise training subsets used by the experiments. The audit does not alter the original partitions, model fitting, hyperparameter selection, or ACC results.

For each of the 18 datasets and each seed from 2 to 11, we examine all six unordered pairs among the four training centers, giving $18 \times 10 \times 6 = 1080$ center pairs and 180 dataset–repetition summaries. For centers $u$ and $v$, let $\mu_{uj}$ and $\mu_{vj}$ be the means of feature $j$, and let $s_j$ be its standard deviation after pooling the four training centers for that dataset and repetition. Here $d'$ denotes the number of nonconstant valid features.

The average absolute standardized mean difference is

$$\mathrm{ASMD}_{uv} = \frac{1}{d'} \sum_{j=1}^{d'} \frac{|\mu_{uj} - \mu_{vj}|}{s_j}. \tag{S1}$$

For the label distribution, if $p_{uc}$ and $p_{vc}$ are the proportions of class $c$ in centers $u$ and $v$, respectively, we use the total variation distance

$$\mathrm{TVD}_{uv} = \frac{1}{2} \sum_{c=1}^{C} |p_{uc} - p_{vc}|. \tag{S2}$$

Across the 180 dataset–repetition summaries, mean pairwise ASMD has a median of 0.0735 (interquartile range 0.0403–0.1011). The dataset-level ten-repetition means range from 0.0274 for Page Blocks to 0.1780 for Sonar Mines/Rocks. Mean pairwise label TVD is much smaller, with a median of 0.00147 (interquartile range 0.00092–0.00210); the largest dataset-level mean is 0.01336 for Wine, while Iris has zero label TVD under the stratified allocation. Figure S4 reports all dataset-level summaries. These values show modest finite-sample feature-mean discrepancies and very small class-proportion discrepancies under the current protocol; they are not predefined shift magnitudes.

Figure S5 provides a complementary visual audit for Sonar Mines/Rocks, Diabetes, and Page Blocks at seed 2. The four center point clouds do not form clear center-exclusive clusters in the displayed PCA planes, and their class-proportion bars are nearly coincident. This pattern is consistent with same-dataset stratified random allocation. However, ASMD describes only feature means, TVD describes only class proportions, and PCA is only a two-dimensional projection. Neither visual overlap nor nonzero ASMD constitutes a formal test of equal distributions, a controlled shift intervention, or evidence of robustness to naturally occurring institutional heterogeneity.

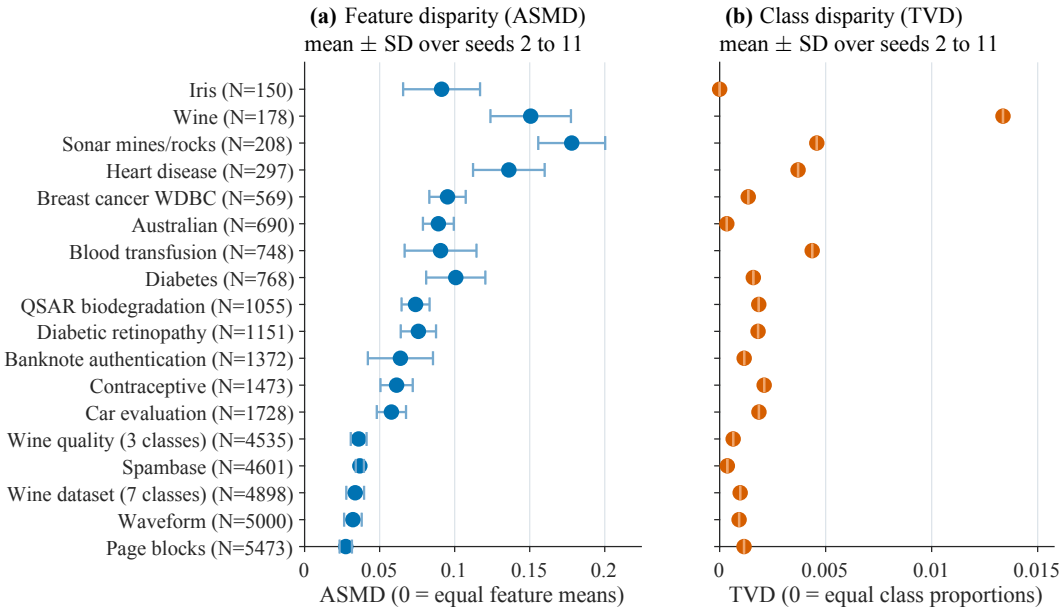


Figure S4: Post-hoc audit of empirical cross-center differences in the existing training splits. Panels (a) and (b) report feature-mean disparity using mean pairwise ASMD and class-proportion disparity using mean pairwise TVD, respectively. Points denote dataset means over seeds 2–11, and error bars denote one standard deviation above and below the mean; each repetition averages the six unordered pairs among the four training centers ($M$, $A_1$, $A_2$, and $T$). The values characterize finite-sample variation under stratified random partitioning; no shift generator was imposed.

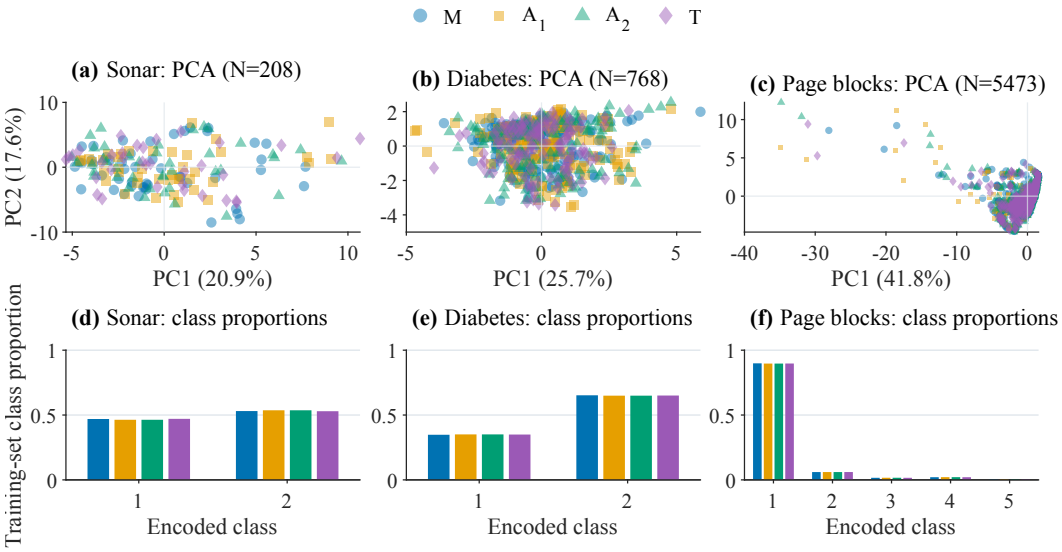


Figure S5: Representative cross-center overlap and class composition at seed 2. Panels (a)–(c) show PCA projections for Sonar Mines/Rocks, Diabetes, and Page Blocks after pooled training-set $z$-standardization; panels (d)–(f) report the corresponding center-wise class proportions. Colors and markers identify the main center ($M$), auxiliary centers ($A_1$ and $A_2$), and target center ($T$). PCA is a descriptive two-dimensional projection and is not a significance test for covariate shift.