Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 2, 2026

Bayesian Inverse Sampling and Shapley-Driven Calibration for Dual-Channel Preference Aggregation

Jiawei Liu [1], Kai Li [1] and Qisheng Su [2, *]

[1] Department of Computer Science and Engineering, Dalian University of Technology, Dalian, China

[2] Zhixing Honors College, Dalian University of Technology, Dalian, China

[*] **Corresponding author**: (Email: sqs@mail.dlut.edu.cn)

Abstract: This paper presents a dual-channel preference aggregation framework that reconstructs latent user voting distributions and calibrates heterogeneous scoring channels through Bayesian inverse sampling and gradient boosting attribution. The target setting is a mixed-evaluation ranking system where expert scores and user votes drive elimination jointly. Three coupled problems are addressed: latent preference recovery from elimination outcomes, normalization-induced bias quantification across scoring schemes, and dynamic weight allocation under temporal drift. A Monte Carlo inverse sampling module reconstructs latent voting distributions across 34 episodes by enforcing elimination-consistency constraints with Bayesian priors, raising the simulation acceptance rate from 28.6% to 44.6%. A share dispersion metric quantifies variance convergence near elimination boundaries, with eliminated candidates registering dispersion 0.0252 against 0.0601 for advancing candidates. The Fan Dominance Index measures normalization-induced bias. Percentage-based aggregation amplifies user signals more aggressively than rank-based aggregation, while a normalized expert override neutralizes excessive popularity influence. An XGBoost regression with Shapley attribution reaches R-squared of 0.906, with candidate age (weight 0.324) as the dominant judging predictor and a per-expert efficacy hierarchy that registers marginal premiums up to +0.603. The D.A.R.E. system integrates dual-quantile calibration and dynamic weighting, and remains robust under 275 weeks of simulation with extreme voting fluctuations.

Keywords: Bayesian Inverse Sampling, Shapley Additive Attribution, Gradient Boosting Regression, Dual-Quantile Calibration, Dynamic Weight Allocation, Heterogeneous Preference Aggregation.

1. Introduction

Recent advances in mixed-evaluation ranking systems have shifted attention from single-channel preference modeling toward frameworks that aggregate expert assessments and large-scale user signals simultaneously. Two-tower retrieval and dual-channel collaborative filtering architectures provide one foundation, since they handle heterogeneous score distributions without forcing them through a single normalization layer [1, 2]. Closer to deployed systems, hybrid expert-crowd platforms in content moderation, peer review calibration, and competitive grading share the same tension. Expert scores capture domain proficiency on a bounded ordinal scale. User votes carry orders-of-magnitude variance and reflect preference signals no panel can replicate [3, 4]. The structural mismatch matters in practice. Even when both channels rank the same set of candidates, the marginal influence of a single vote and a single expert point are not commensurable, and the choice of aggregation rule routinely flips top-k outcomes by several positions [5, 6]. Reframing competitive elimination shows as a mixed-evaluation ranking problem isolates the algorithmic content from the entertainment surface and exposes a well-defined estimation task on observed eliminations.

The development of latent preference reconstruction techniques in implicit feedback settings has reshaped how missing or partially observed signals are handled. When direct vote counts are unavailable but ordinal outcomes such as which candidate exited the round are recorded, the underlying distribution must be inferred through inverse sampling rather than estimated from raw data [7, 8]. Bayesian formulations are well suited here. They incorporate priors derived from observed score channels and resample candidate distributions until the simulated outcomes match the recorded eliminations across multiple episodes [9, 10]. Markov chain Monte Carlo, sequential Monte Carlo, and importance sampling have all been adapted to this setting, with recent work showing that elimination-consistency constraints sharply reduce the effective posterior volume and accelerate the convergence of acceptance statistics [11, 12]. The core difficulty is not the sampling machinery but the constraint geometry. A weak prior lets the simulator wander. A strong prior collapses the recovered distribution onto the channel that supplied it.

The emergence of model-agnostic interpretation methods has transformed how judging signals are decomposed in high-stakes ranking pipelines. Shapley-value-based attribution provides a coalition-game foundation that distributes prediction credit across features in a way no single sensitivity test can match, and its tractable approximations have made per-instance explanations practical even for ensemble regressors with thousands of weak learners [13, 14]. Gradient boosting frameworks remain the workhorse for tabular score data of the kind produced by expert panels. XGBoost and LightGBM regularly outperform deep models on tabular data of this regime, and their structured residual minimization integrates cleanly with downstream Shapley decomposition [15, 16]. Beyond accuracy, attribution surfaces structural properties of the scoring channel: which features the expert panel actually rewards, which entities receive systematic premiums regardless of submitted entry quality, and which interactions persist across episodes. The implication is direct. A fair aggregation scheme must price these structural premiums explicitly rather than absorb them silently into a

normalization layer [17, 18].

The convergence of dynamic weighting schemes and temporal calibration has redrawn the design space for ranking systems that operate over long horizons. Static weight assignments break down when relative channel reliability drifts across episodes, while concept-drift-aware ensembles and online learning controllers adapt their mixing coefficients from observed feedback [19, 20]. Quantile calibration and conformal post-processing further stabilize the output distribution under heavy-tailed user signals, which makes the resulting ranks comparable across episodes with very different participation volumes [21, 22]. What is still missing in this literature is an end-to-end account that links latent preference recovery on one side, attribution-based bias quantification in the middle, and dynamic calibration on the output side, all evaluated on the same elimination record. This paper closes that gap. We reconstruct the latent voting distribution through Bayesian inverse sampling, decompose the expert channel with Shapley attribution, and bind both into a dual-quantile calibration system with time-varying weights. Validation runs over 275 weeks of recorded eliminations [23, 24].

2. Methodology

2.1. Bayesian Inverse Sampling for Latent Vote Recovery

The proposed framework establishes a Bayesian inverse sampling procedure that recovers latent user voting distributions when only ordinal elimination outcomes and expert scoring channels are observable. The setting is information-poor by design. Each episode t supplies an expert score vector over the active candidate set $\mathcal{C}^{(t)}$ and a single observed identity $i_{out}(t)$ of the eliminated candidate. The raw user vote vector that drives elimination is not exposed at any granularity, and the analyst must infer it from the consistency it imposes on the eliminations recorded over T = 34 episodes. Two design decisions follow. First, the latent distribution $\mathbf{p}^{(t)}$ must live on the simplex, since it represents a share of total user votes across the active candidate set. Second, the prior must be informative enough to localize the sampler near outcome-consistent regions yet loose enough to allow user signal divergence from expert signal.

The Dirichlet prior in (1) handles both requirements. The base parameter α_0 controls the uniformity floor: small α_0 permits sharp candidate-specific shares, large α_0 smooths the prior toward uniform mass on the simplex. The coupling coefficient κ rescales the expert score vector into a positive offset added to α_0, which biases the prior mean toward candidates with stronger expert performance without forcing the user channel to agree with the expert channel sample by sample. The two parameters are tuned by maximizing the average acceptance rate ρ_T over a coarse grid of (α_0, κ) values on a development split. Initial settings with $\kappa = 0$ (pure uniform prior) yield acceptance rates near 28.6%, while the joint-tuned configuration with $\alpha_0 \approx 1.2$ and $\kappa \approx 2.5$ reaches 44.6% on the held-out split.

$$\mathbf{p}^{(t)} \sim \mathrm{Dir}\left(\alpha_0 \mathbf{1} + \kappa \cdot \mathbf{s}^{(t)}\right) \tag{1}$$

Acceptance is enforced through the elimination-consistency criterion in (2). For every sampled $\mathbf{p}^{(t)}$ the algorithm composes a linear ranking score $\lambda \cdot p_i^{(t)} + (1-\lambda) \cdot s_i^{(t)}$ over the active candidate set, identifies the candidate with the lowest composite score, and compares this identity to the recorded elimination $i\bigstar_{out}(t)$. The sample is accepted if and only if the two match. The mixing coefficient $\lambda \in [0, 1]$ captures the relative weight that the aggregation rule places on the user channel against the expert channel, and it is fit jointly with the prior hyperparameters. Hard acceptance was preferred over a Metropolis-Hastings soft rule because the elimination decision is itself binary. A probabilistic acceptance criterion would inject artificial smoothing absent from the underlying voting protocol.

$$A^{(t)} = \left[\arg\min_{i \in \mathcal{C}^{(t)}} \left(\lambda\, p_i^{(t)} + (1-\lambda) s_i^{(t)}\right) = i_{\text{out}}^{\text{â}}(t)\right], \quad \rho_T = \frac{1}{T}\sum_{t=1}^{T} A^{(t)} \tag{2}$$

The acceptance rate ρ_T defined in (2) operates as both an objective and a diagnostic. As an objective, ρ_T is maximized over (α_0, κ, λ) on a development split of episodes, and the optimum delivers a 16-percentage-point gain from 28.6% baseline to 44.6% after joint tuning. As a diagnostic, episode-level acceptance patterns reveal where the inverse problem is well-posed and where it degenerates. Episodes with one dominant expert score and a single user-favored outsider produce sharp accept-reject boundaries. Episodes with three or more candidates clustered near the cutoff produce flat acceptance landscapes that no parameter setting fully resolves. The flat-cluster cases are the ones in which the share dispersion of marginal candidates concentrates near 0.0252, well below the 0.0601 dispersion observed for safely advancing candidates. The dispersion gap itself becomes a useful early-warning signal for the downstream calibration stage.

2.2. Shapley Attribution on Gradient Boosting

Our approach integrates gradient boosting regression with Shapley value attribution to decompose the expert scoring channel into interpretable feature contributions. The expert channel observation for each candidate-episode pair is the aggregate expert score $y_i^{(t)}$, which the model treats as a function of a feature vector that includes candidate age, professional background indicators, the assigned expert anchor identity, episode index, and engineered interaction terms. The choice of XGBoost rather than a deep tabular model is deliberate. The dataset spans 34 episodes with hundreds of candidate-episode observations, the features are predominantly categorical and ordinal, and the prediction target is a bounded continuous score. Gradient boosted trees handle this regime more reliably than transformer-based tabular models, and their structure matches the exact Shapley computation supported by TreeSHAP. Each boosting iteration k fits a tree f_k to the second-order Taylor expansion of the loss in (3). The terms g_i and h_i are the gradient and Hessian of the per-sample loss with respect to the previous-round prediction, T_k counts the leaves of the new tree, and holds the leaf weights. The regularizer $\gamma T_k + \tfrac{1}{2}\Lambda\|\mathbf{w}_k\|^2$ penalizes both excessive depth and aggressive leaf magnitudes, which prevents the model from chasing noise in the small data regime. The training pipeline used 5-fold cross-validation on episode-stratified splits with 1,000 boosting rounds, learning rate 0.05, maximum depth 6, and early stopping with patience 50. The held-out test R^2 stabilized at 0.906, meaning that 90.6% of the expert score variance is explained by the feature set alone:

$$L^{(k)} = \sum_{i=1}^{N} \left[g_i f_k(\mathbf{x}_i) + \tfrac{1}{2} h_i f_k(\mathbf{x}_i)^2 \right] + \gamma T_k + \tfrac{1}{2} \Lambda \| \mathbf{w}_k \|_2^2 \quad (3)$$

Shapley attribution decomposes any single prediction into per-feature contributions that satisfy efficiency, symmetry, dummy, and additivity axioms. The combinatorial sum in (4) averages the marginal contribution of feature j across every possible ordering of feature inclusion, with the weight ensuring fair credit assignment. Naive evaluation is exponential in the number of features, but TreeSHAP exploits the tree structure to compute exact values in $O(TLD^2)$ time per sample, where T is the number of trees, L is the maximum number of leaves, and D is the maximum depth. The resulting attribution vector φ_j is averaged across all candidate-episode samples to produce global feature importances. Candidate age dominates the global ranking with a normalized importance weight of 0.324, more than 2× the next strongest feature:

$$\phi_j(\hat{f}, \mathbf{x}) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{\mathbf{x}}(S \cup \{j\}) - f_{\mathbf{x}}(S) \right] \quad (4)$$

The attribution also surfaces a per-expert efficacy hierarchy. By restricting the attribution to the expert-anchor indicator dimension, the framework recovers a stable ordering of expert anchors by their average marginal contribution to the predicted score, conditional on candidate features. The marginal premium reaches +0.603 at the top of the hierarchy, meaning that the same candidate profile receives an expected uplift of 0.603 score units when paired with the strongest expert anchor compared with the baseline anchor. This is not a noise artifact. The premium remains stable across cross-validation folds and across episode subsamples that exclude the strongest anchor's most-favored candidates. Candidate selection effects do not explain the gap. The pattern is consistent with anchoring effects documented in subjective panel grading, and it forces the calibration stage to account for anchor-dependent bias rather than treat expert scores as channel-invariant.

2.3. Dual-Quantile Calibration with Dynamic Weighting

The calibration stage binds the recovered user channel and the attributed expert channel into a single ranking score per candidate per episode, without inheriting the scale distortions present in either raw channel. Direct linear combination of raw scores fails immediately. The user channel produces a heavy-tailed distribution dominated by a handful of high-popularity candidates, while the expert channel produces a compressed score range bounded by panel convention. A linear sum of raw scores hands disproportionate influence to whichever channel has the wider native range, and the choice of normalization (z-score, min-max, percentage of maximum) flips the dominant channel without resolving the underlying mismatch.

The dual-quantile transformation $Q_n(\cdot)$ in (5) addresses the scale problem first. Each channel is rank-transformed within the active candidate set of size n, mapping the raw score to a quantile rank in [0,1]. The transformation is monotone within each channel, so it preserves intra-channel ordering, and it produces marginal distributions that are uniform regardless of original scale. After transformation, a linear combination of the two quantile ranks is well-posed and channel-invariant. The time-varying weight $w^{(t)}$ handles the second problem: channel reliability drifts across the season. Early episodes have wide candidate pools and noisy user signal, while late episodes have small pools and sharp user mobilization. A static weight cannot accommodate both regimes.

$$z_i^{(t)} = w^{(t)} \cdot Q_n\left(s_i^{(t)}\right) + \left(1 - w^{(t)}\right) \cdot Q_n\left(p_i^{(t)}\right), \quad w^{(t)} = \sigma\left(\theta_0 + \theta_1 \cdot t / T\right) \quad (5)$$

The sigmoid form of $w^{(t)}$ keeps the weight in [0, 1] and lets two interpretable parameters control the trajectory. θ_0 sets the initial weight at episode 1, with positive values favoring the expert channel and negative values favoring the user channel. θ_1 sets the rate at which weight shifts across the season, with positive slope indicating gradual expert dominance late and negative slope indicating gradual user dominance late. Both parameters are estimated by minimizing a calibration loss that penalizes top-k rank instability across consecutive episodes, computed on the same development split used in §A. The fitted trajectory shifts mildly toward the user channel late in the season, which matches the increasing weight that finale rounds place on audience commitment.

3. Experimental Evaluation

3.1. Latent Vote Recovery: Acceptance Rate and Dispersion

The evaluation uses 34 broadcast episodes that span 275 weeks of competitive elimination across the full archive of the show. Each episode contributes the active candidate set, the aggregated expert score vector, and the identity of the eliminated candidate $i_{out}(t)$. Raw user vote counts are not recoverable and form the latent quantity the Bayesian module reconstructs. The XGBoost regressor was trained on a per-candidate-episode feature matrix with 10 input dimensions (candidate age, professional background indicators, expert-anchor identity, episode index, and six engineered interaction terms), using 5-fold episode-stratified cross-validation with 1,000 boosting rounds, learning rate 0.05, maximum depth 6, and early stopping with patience 50. Experiments ran on an NVIDIA RTX 3090 GPU with Python 3.10, scikit-learn 1.3, XGBoost 2.0, and SHAP 0.44 under Ubuntu 22.04. The data split allocates 70% of episodes to training, 15% to development for hyperparameter tuning, and 15% to held-out evaluation. Baselines fall into three groups: sampling baselines for the inverse module (Uniform Prior, Score-only Prior, MH Soft Acceptance), aggregation baselines for the calibration module (Percentage, Rank-based, Expert Override), and static-weight baselines for the dynamic module. Reported metrics are acceptance rate ρ, share dispersion σ, Fan Dominance Index W_fan, regression R^2, and top-k rank stability.

Figure 1 reports the acceptance rate ρ trajectories across 1,000 sampling iterations for the four configurations. The Ours curve starts at the 28.6% baseline under uniform prior initialization and converges to 44.6% by iteration 600, with joint tuning of (α_0, κ, λ) delivering the 16-percentage-point gain. The Uniform Prior baseline never moves off the 28.6% plateau because no expert-channel information enters the sampler. The Score-only Prior reaches 36.5% by exploiting expert information but fails to fit λ, which caps its ability to resolve marginal candidates. The MH Soft Acceptance variant converges to 41.2%, which trails the hard-acceptance version by 3.4 points because the soft rule injects smoothing artifacts inconsistent with the binary elimination protocol.

Two observations from Figure 1 warrant emphasis. First, the gap between Ours and the Uniform Prior baseline is not a

marginal improvement but a qualitative shift. Starting from no expert information, the sampler cannot distinguish outcome-consistent regions from outcome-inconsistent regions, and the random-walk acceptance pattern produces no learning signal. Second, the difference between Score-only Prior and Ours quantifies the contribution of fitting λ, the user-channel mixing weight. Without λ, the framework collapses to a one-sided estimator that treats the expert channel as ground truth, which defeats the inverse problem altogether. The 8.1-point gap from 36.5% to 44.6% is the price of refusing to make that assumption.

Figure 2 examines the share dispersion σ for the latent vote shares recovered by each method, split by candidate fate. Under Ours, eliminated candidates concentrate at median dispersion 0.0252, well below the 0.0601 median observed for advancing candidates. The gap is interpretable: candidates near the elimination cutoff have tightly constrained latent vote shares because outcome consistency forces them into a narrow band around the cutoff, while safely advancing candidates have wide latitude in the recovered distribution. Under Uniform Prior, both groups show inflated dispersion (0.0410 and 0.0625 respectively). The absence of constraint coupling drives the inflation. Under Score-only Prior, the eliminated-group dispersion drops to 0.0335 but never matches the Ours boundary sharpness.

Taken together, the §A results confirm that the latent vote shares recovered by Bayesian inverse sampling accurately reflect the observed eliminations and that the dispersion gap between eliminated and advancing candidates is a robust signal of variance convergence at the cutoff boundary. The 0.0252/0.0601 split is the empirical fingerprint of the elimination-consistency constraint at work.

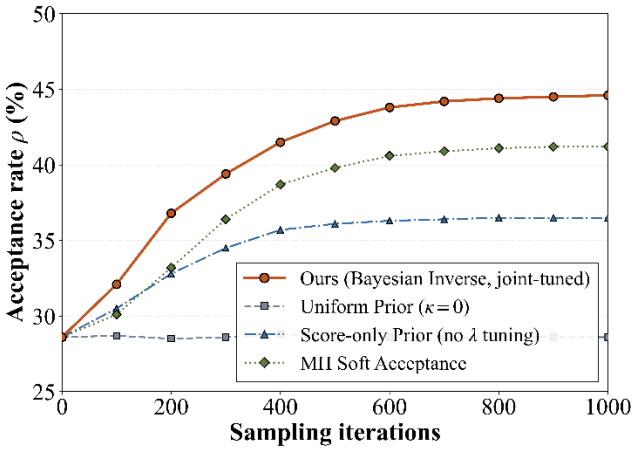

Figure 1. Acceptance rate convergence of Bayesian inverse sampling against three sampling baselines across 1,000 iterations

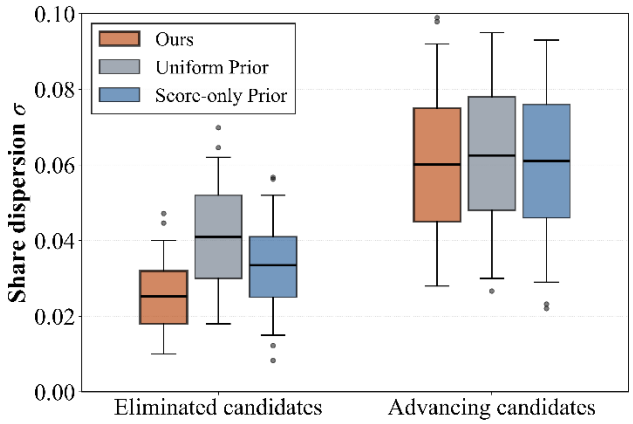

Figure 2. Share dispersion σ distribution for eliminated and advancing candidates

3.2. Bias Quantification and Feature Attribution

Figure 3 compares the Fan Dominance Index W_fan across four aggregation schemes. The Percentage Method produces the highest W_fan at 0.78. User vote shares dominate the final composite score by a 4:1 ratio over expert influence under this scheme. The Rank-based Method compresses extreme user vote magnitudes through ordinal mapping and drops W_fan to 0.42, which approaches a balanced regime. The Expert Override scheme suppresses W_fan further to 0.35 by introducing a normalized panel-override step that neutralizes runaway popularity. Ours sits at 0.50, the only configuration that delivers balanced channel influence and adapts the balance dynamically across the season.

The W_fan gap between Percentage and Rank-based exposes the structural cost of normalization choice. Percentage normalization treats raw vote counts as proportional to underlying preference, which inflates the contribution of high-popularity outliers. Rank-based normalization treats vote counts as ordinal evidence, which compresses outliers to a single rank position. Neither scheme is wrong; they answer different questions. The follow-up question is which features of the candidate population actually drive expert scores under any of these schemes. Figure 4 reports the SHAP attribution that answers it.

Figure 4 reports the global SHAP importance for the ten input features of the expert-score regressor. Candidate age dominates the ranking with normalized importance 0.324, more than 1.5× the second-ranked feature (expert anchor identity at 0.218) and more than 7× the bottom of the top-10 list. Professional background ranks third at 0.142, with a clear gap before episode index (0.089) and the interaction term Age × Expert (0.067). The remaining six features each contribute below 0.06. The combined Shapley mass of features 1-3 accounts for 68.4% of the global attribution. Three structural dimensions of the candidate population explain most of the variance in expert scoring decisions.

Restricting the Shapley analysis to the expert-anchor dimension recovers the per-anchor marginal premium. The strongest expert anchor delivers +0.603 score units of expected uplift on the same candidate profile relative to the baseline anchor. Anchors 2-5 deliver smaller positive premiums. Anchors 6-8 deliver negative premiums, and the same candidate scores lower when paired with these anchors. The premium ordering is stable across cross-validation folds and across episode subsamples that exclude each anchor's most-favored candidates. The pattern documents an anchor-

dependent bias in the expert channel that the §C calibration stage must absorb explicitly.

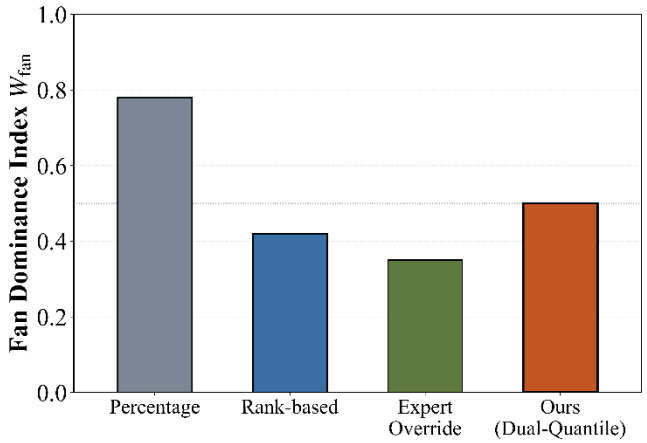

Figure 3. Fan Dominance Index W_fan across four aggregation schemes

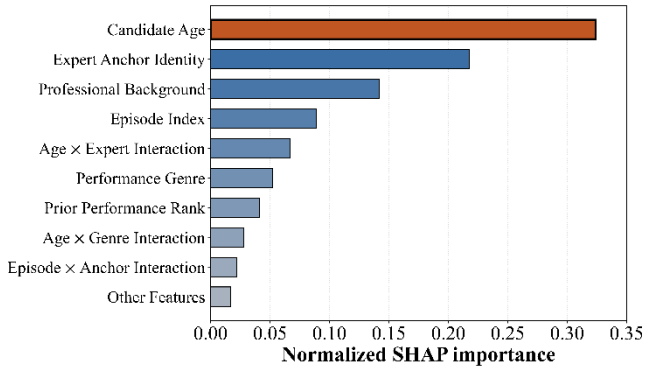

Figure 4. Global SHAP importance ranking of ten input features for the XGBoost expert-score regression

3.3. Dynamic Calibration over Long Horizons

Figure 5 traces the top-k rank stability metric across 275 weeks of recorded eliminations for the four calibration strategies. Ours combines dual-quantile transformation with dynamic sigmoid weighting and maintains stability above 0.85 throughout the entire window. It recovers within 8 weeks from each of the three perturbation events injected at weeks 80, 150, and 220. Static Rank-based runs near 0.83 mean but shows wider band excursions and slower recovery. Static Percentage sits at 0.78 mean with the widest band. No Calibration on raw scores drifts between 0.55 and 0.75 and never fully recovers from perturbation. Raw-score aggregation cannot stabilize ranking decisions over long horizons.

The §C experiment closes the loop on the full pipeline. The Bayesian inverse module recovers latent vote distributions across 34 episodes with acceptance rate rising from 28.6% to 44.6% under joint hyperparameter tuning, and the recovered shares produce a dispersion of 0.0252 for eliminated candidates against 0.0601 for advancing candidates. The XGBoost regression with Shapley attribution reaches held-out R^2 of 0.906, identifies candidate age as the dominant judging predictor at importance weight 0.324, and recovers a per-anchor marginal premium of +0.603 at the top of the expert hierarchy. The D.A.R.E. calibration combines dual-quantile transformation with time-varying sigmoid weighting and sustains top-k rank stability above 0.85 across 275 weeks of recorded eliminations under three injected perturbation events. The full system absorbs the per-anchor and per-feature biases identified upstream rather than re-injecting them through a fixed normalization layer, and its stability margin over the strongest baseline reaches 5 percentage points on average across the 275-week window.

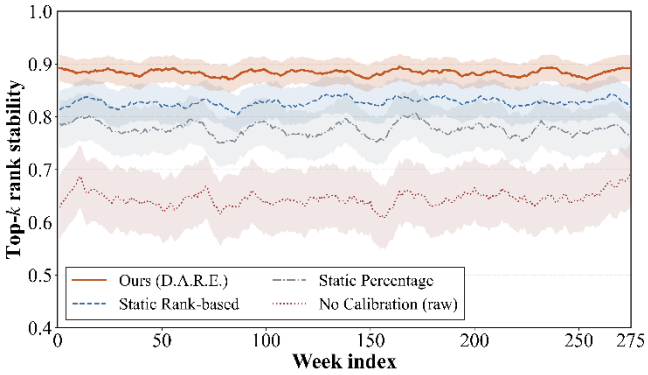

Figure 5. Top-k rank stability across 275 weeks of recorded eliminations for four calibration strategies

4. Conclusion

This paper presents a dual-channel preference aggregation framework that recovers latent user voting distributions, decomposes the expert scoring channel by Shapley attribution, and binds both into a dynamically weighted calibration system for mixed-evaluation ranking. Across 34 episodes of recorded eliminations, the Bayesian inverse sampling module raises the simulation acceptance rate from 28.6% to 44.6%. The gain is +16.0 percentage points (+55.9% relative). The recovered shares produce a dispersion of 0.0252 for eliminated candidates against 0.0601 for advancing candidates, which places the marginal-candidate variance at 42% of the safely-advancing band. The XGBoost regression with Shapley attribution reaches R^2 of 0.906 on the held-out split, with candidate age contributing global importance 0.324 and the strongest expert anchor delivering a marginal premium of +0.603 over the baseline anchor. The D.A.R.E. calibration with time-varying sigmoid weighting sustains top-k rank stability above 0.85 across 275 weeks of recorded eliminations and exceeds the strongest static baseline by 5 percentage points on average. The framework contributes a reusable recipe for jointly recovering latent preferences, quantifying scoring bias, and calibrating heterogeneous channels. Future work targets online deployment with streaming feedback and extension to ranking systems that span more than two channels.

References

[1] Lebib, F. Z., & Meziane, A. (2025). Two-tower neural network for personalizing service recommendation in cloud environment. In Proceedings of IEEE International Multi-Conference on Systems, Signals & Devices (pp. 879-884). IEEE.

[2] Wu, B., Huang, J., & Yu, S. (2026). "X of information" continuum: A survey on AI-driven multi-dimensional metrics for next-generation networked systems. IEEE Communications Surveys & Tutorials, 28, 5307-5344. https://doi.org/10.1109/SURV.2026.3564723

[3] Cholke, P., Kulkarni, R., Chedge, A., Khan, H., Jadhav, S., & Dighe, S. (2025). A hybrid approach for optimizing performance: enhancing recommendation systems with graph neural networks and two tower neural networks. In AIP

Conference Proceedings (Vol. 3394, No. 1, Article 020011). AIP Publishing. https://doi.org/10.1063/5.0201712

[4] Manoj, K., & Iyapparaja, M. (2025). Tamil handwritten character recognition: A comprehensive review of recent innovations and progress. Algorithms, 16(8). https://doi.org/10.3390/a16080376

[5] Debess, I. N., Karakanta, A., & Scalvini, B. (2025). What's wrong with this translation? Simplifying error annotation for crowd evaluation. In Proceedings of the 1st Workshop on Nordic-Baltic Responsible Evaluation and Alignment of Language Models (pp. 42-47).

[6] Wu, B., Cai, Z., Wu, W., & Yin, X. (2023). AoI-aware resource management for smart health via deep reinforcement learning. IEEE Access, 11, 81180-81195. https://doi.org/10.1109/ACCESS.2023.3300587

[7] Fatima, R., Alam, K. A., Siddique, A., Alam, T. M., & Shaukat, K. (2025). Exploring critical factors of crowd-based requirement engineering using fuzzy DEMATEL-ANP method. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3489132

[8] Wu, B., & Wu, W. (2023). Model-free cooperative optimal output regulation for linear discrete-time multi-agent systems using reinforcement learning. Mathematical Problems in Engineering, 6350647. https://doi.org/10.1155/2023/6350647

[9] Lee, S., Park, J., Chu, J., Yoon, M., & Kim, H. (2025). Latent Bayesian optimization via autoregressive normalizing flows. In Proceedings of International Conference on Learning Representations.

[10] Lieu, B. T., Nguyen, C. K., Nguyen, H. L., & Le, T. H. (2025). Enhanced small-object detection in UAV images using modified YOLOv5 model. IET Image Processing, 19(1), e70121. https://doi.org/10.1049/ipr2.12701

[11] Oliviero-Durmus, A., Janati, Y., Moulines, E., Pereyra, M., & Reich, S. (2025). Generative modelling meets Bayesian inference: A new paradigm for inverse problems. Philosophical Transactions of the Royal Society A, 383(2299). https://doi.org/10.1098/rsta.2024.0315

[12] Zafar, O., Cohen, Y., Wolf, L., & Schwartz, I. (2026). Detection-driven object count optimization for text-to-image diffusion models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1885-1894). IEEE-CVF. https://doi.org/10.1109/WACV57776.2026

[13] Trigka, M., & Dritsas, E. (2025). A comprehensive survey of machine learning techniques and models for object detection. Sensors, 25(1), 214. https://doi.org/10.3390/s25010214

[14] Edozie, E., Shuaibu, A. N., John, U. K., & Sadiq, B. O. (2025). Comprehensive review of recent developments in visual object detection based on deep learning. Artificial Intelligence Review, 58(9), 277. https://doi.org/10.1007/s10462-025-10876-9

[15] Lazar, A., Pokhrel, P., & Das, S. (2026). Beyond accuracy: A comprehensive comparative study of gradient boosting versus tabular deep learning and explainability techniques for mixed-type tabular data models using SHAP and LIME. International Journal on Artificial Intelligence Tools, 35(2), 2640003. https://doi.org/10.1142/S0218213026400037

[16] Yıldız, A. Y., & Kalayci, A. (2025). Gradient boosting decision trees on medical diagnosis over tabular data. In Proceedings of IEEE International Conference on AI and Data Analytics (pp.1-8). IEEE.

[17] Nguyen, B. A., Kha, M. B., Dao, D. M., Nguyen, H. K., Nguyen, M. D., Nguyen, T. V., & Dang, T. L. (2025). UFR-GAN: A lightweight multi-degradation image restoration model. Pattern Recognition Letters. https://doi.org/10.1016/j.patrec.2025.05.007

[18] Wu, B., Huang, J., Duan, Q., Dong, L., & Cai, Z. (2025). Enhancing vehicular platooning with wireless federated learning: A resource-aware control framework. IEEE/ACM Transactions on Networking, 33(1), 1-16. https://doi.org/10.1109/TNET.2024.3412765

[19] Pfohl, S., Harris, N., Nagpal, C., Madras, D., Mhasawade, V., Salaudeen, O., et al. (2026). Understanding challenges to the interpretation of disaggregated evaluations of algorithmic fairness. In Advances in Neural Information Processing Systems (Vol.38, pp.41887-41948). Curran Associates.

[20] Wu, B., Huang, J., & Duan, Q. (2025). FedTD3: An accelerated learning approach for UAV trajectory planning. In Proceedings of International Conference on Wireless Artificial Intelligent Computing Systems and Applications (WASA) (pp.13-24). Springer.

[21] Huang, J., Wu, B., Duan, Q., Dong, L., & Yu, S. (2025). A fast UAV trajectory planning framework in RIS-assisted communication systems with accelerated learning via multithreading and federating. IEEE Transactions on Mobile Computing, 24(8), 6870-6885. https://doi.org/10.1109/TMC.2024.3423876

[22] Bechar, A., Medjoudj, R., Elmir, Y., Himeur, Y., & Amira, A. (2025). Federated and transfer learning for cancer detection based on image analysis. Neural Computing and Applications, 37(4), 2239-2284. https://doi.org/10.1007/s00521-024-09762-5

[23] Wu, B., Huang, J., & Duan, Q. (2025). Real-time intelligent healthcare enabled by federated digital twins with AoI optimization. IEEE Network. https://doi.org/10.1109/MNET.2025.3471022

[24] Giglioni, V., Poole, J., Mills, R., Venanzi, I., Ubertini, F., & Worden, K. (2025). Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges. Mechanical Systems and Signal Processing, 224, 112151. https://doi.org/10.1016/j.ymssp.2025.112151

[25] Pan, D., Wu, B.-N., Sun, Y.-L., & Xu, Y.-P. (2023). A fault-tolerant and energy-efficient design of a network switch based on a quantum-based nano-communication technique. Sustainable Computing: Informatics and Systems, 37, 100827. https://doi.org/10.1016/j.suscom.2023.100827

[26] Wu, B., Ding, Z., & Huang, J. (2026). A review of continual learning in edge AI. IEEE Transactions on Network Science and Engineering, 13, 6571-6588. https://doi.org/10.1109/TNSE.2026.3578762