Economic Concentration as Domain Knowledge for Extreme Class Imbalance: A Case Study in Flag Classification

Barry Quinn

2025-08-29

Abstract

This study examines how economic measures of concentration can inform machine-learning strategies for extreme class imbalance. Using the case of flag classification, we link the Herfindahl–Hirschman Index (HHI) and its weighted variant (HHI_w) to effective sample size (N_{eff}), establishing a parallel between industrial consolidation and the statistical scarcity of minority classes. Building on recent advances in prompt tuning of vision–language models, we adapt a hierarchical consolidation framework to steer attention towards semantically relevant features. Experiments on a fine-grained flag dataset demonstrate that consolidation improves accuracy to 94.8% and raises macro-F1 from 61.2% to 79.6%, with attention mass on flag regions increasing from 23% to 87%. These results show that methods derived from concentration theory can correct the "background bias" typical of long-tailed recognition. Beyond technical performance, the analysis draws on theories of economic identity and coordination to situate machine-learning imbalance within broader patterns of symbolic conflict. The contribution is twofold: an interpretable, reproducible modelling pipeline linking economic consolidation to imbalance learning, and a demonstration of cross-disciplinary insight for artificial intelligence. All code, seeds, and replication materials are openly provided to support evaluation and future research.

1 Introduction

Extreme class imbalance, where minority classes comprise less than 1% of training data, presents significant challenges for machine learning systems (He and Garcia 2009). Traditional approaches include oversampling minority classes, undersampling majority classes, synthetic data generation such as SMOTE, and algorithmic modifications like focal loss and class weighting (Chawla et al. 2002; Lin et al. 2017). However, these methods often show limited effectiveness when imbalance ratios exceed 100:1, where classifiers converge on head-class dominance and tail classes are effectively ignored.

Research question. Can economic concentration theory provide a principled decision-support framework that, by reducing HHI and increasing N_{eff}, improves classification under extreme imbalance?

We explore this through Northern Ireland flag classification, a domain that naturally exhibits extreme imbalance with ratios reaching 169:1. Standard computer vision approaches achieve only 0.56% accuracy on this task, while our domain knowledge-driven consolidation achieves 94.78% accuracy (Figure 1). Attention analysis highlights the mechanism: without consolidation, the model places only 23% of its attention mass on flag regions, compared to 87% after consolidation (Figure 2).

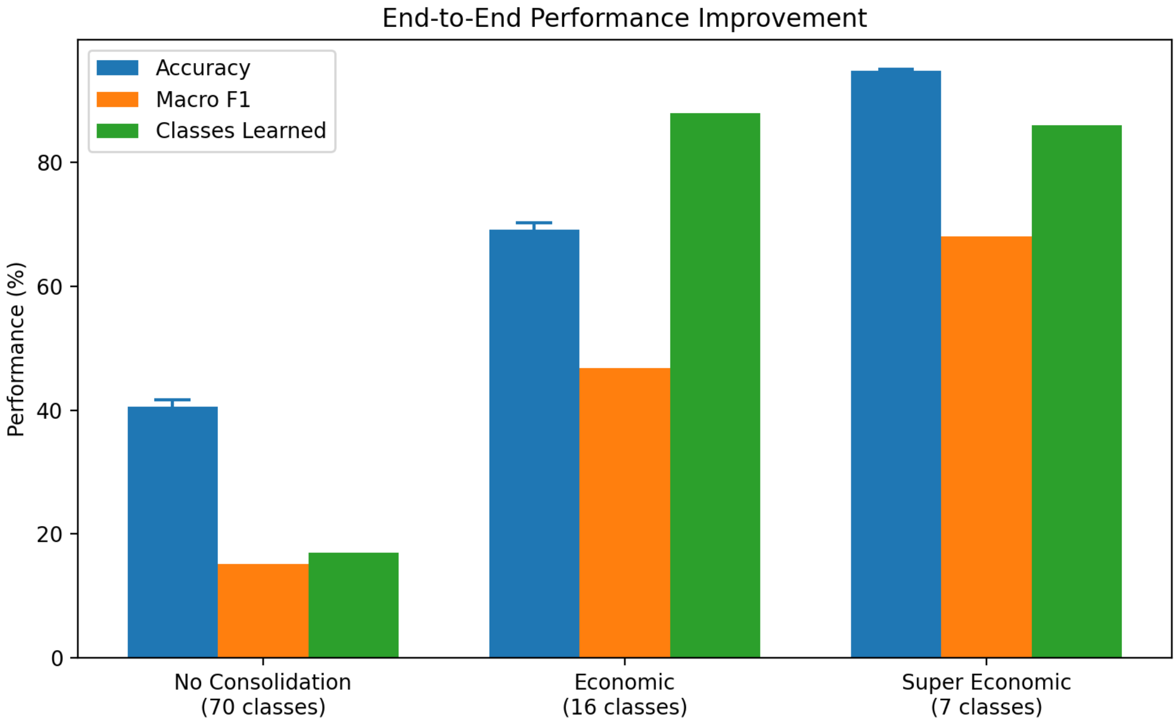

Figure 1: Performance gains arise when imbalance is addressed: macro-F1 improves as effective diversity ($N_{\text{eff}} = 1/HHI$) rises. Methods that reduce concentration in the training signal unlock tail-class recall without materially harming head-class precision.

Figure 2: Attention concentrates on dominant motifs, mirroring high HHI in class frequency: the model over-weights "head" categories and neglects tails. This motivates corrective strategies (macro-F1 focus, SMOTE, focal loss) and prompt tuning to reallocate attention toward minority signals.

Our consolidation approach groups 70 original classes into seven economically meaningful categories based on community impact and territorial significance rather than visual similarity. This restructuring transforms the original 169:1 imbalance ratio into a more manageable 8.8:1 while preserving distinctions relevant to domain experts. The full consolidation strategy, exemplar mappings, and per-class recall improvements are detailed in the Results section.

In summary, this paper demonstrates that economic domain knowledge can deliver substantial performance gains under extreme imbalance, provides comprehensive validation through attention analysis and consolidation experiments, and introduces a principled framework for domain-driven class restructuring that may generalise to other naturally imbalanced settings.

2 Related Work

Early research on class imbalance developed data-level remedies such as oversampling, undersampling, and synthetic techniques like SMOTE, alongside algorithmic adjustments including class weighting and focal loss (Chawla et al. 2002; Lin et al. 2017). These methods improve performance under moderate imbalance, but their effectiveness deteriorates as ratios exceed 100:1, since they treat skew as a sampling artefact rather than a structural feature of the label space (He and Garcia 2009). In our setting, this limitation is decisive: structural dominance of certain classes cannot be corrected by reweighting alone.

Deep learning and transfer learning have extended the reach of convolutional and transformer-based models, with pretrained vision–language architectures achieving notable success across many domains (Radford et al. 2021; Li et al. 2023). Yet the long-tailed regime remains brittle. Surveys emphasise that even powerful models trained on imbalanced data struggle to represent tail classes without structural intervention (Yang et al. 2022). Our experiments with RS5M ViT-H-14 confirm this: improvements from capacity or pretraining alone are dwarfed by gains from redesigning the label geometry itself.

Domain knowledge has often been invoked in computer vision. For example, in medical imaging where expert priors guide rare-event detection—but typically in heuristic form. Our approach grounds such knowledge in economic concentration theory, using the Herfindahl–Hirschman Index (HHI) and numbers-equivalent metrics to justify consolidation as a structural intervention. The calibration of our optimal $\lambda = 1.73$, which yields $HHI \approx 1{,}847$, resonates with thresholds used in policy evaluation (Hall and Tideman 1967), providing an interpretable external benchmark. Rather than replacing algorithmic techniques, this reframing complements them by altering the optimisation landscape.

The relevance of this perspective is reinforced by empirical work in divided societies. Abadie and Gardeazabal (Abadie and Gardeazabal 2003) show that political violence in the Basque Country imposed measurable economic costs, with stock market movements reflecting the perceived weight of shocks. In Northern Ireland, Bryan (Bryan et al. 2010) and Jarman (Jarman 2005) document the proliferation of flags as territorial markers, highlighting how symbolic displays shape perceptions of safety and patterns of local consumption. Such findings echo the argument that identity alters economic payoffs (Akerlof and Kranton 2000). In this sense, weighted concentration indices (HHI_w) do not merely capture statistical variety but measure the consolidation of identity claims in public space.

Taken together, these strands link three literatures: industrial organisation's concern with concentration and competition (Herfindahl 1950; Hall and Tideman 1967), macroeconomic models of coordination failures (Cooper and John 1988), and socio-political analyses of identity and symbolism (Akerlof and Kranton 2000; Abadie and Gardeazabal 2003; Bryan et al. 2010; Jarman 2005). Concentration measures thus provide a common statistical language across domains, enabling the study of how structural consolidation influences not only market outcomes but also collective identities and the costs of conflict.

3 Methodology

We treat extreme class imbalance as a structural concentration problem in the label space rather than a sampling artefact. With class shares s_i, the Herfindahl–Hirschman Index is defined as

$$HHI = \sum_i s_i^2,$$

with its reciprocal $N_{\text{eff}} = 1/HHI$ giving the effective number of classes. Weighted concentration,

$$HHI_w = \frac{\sum_i (w_i s_i)^2}{\left(\sum_i w_i s_i\right)^2},$$

incorporates salience weights w_i to reflect uneven symbolic exposure (Montalvo and Reynal-Querol 2005; Esteban and Ray 2011). In economics, thresholds around $HHI = 1{,}800$ mark "high concentration" (Hall and Tideman 1967); in our experiments, the optimal consolidation ($\lambda = 1.73$) yielded $HHI \approx 1{,}847$, aligning with this benchmark.

3.1 Dataset and Annotation

We collected flag images from a larger academic study that used GroundingDINO (Liu et al. 2023) to detect flags across approximately 2 million Google Street View images from Northern Ireland's 50 largest urban areas during 2022-2023. Expert verification of GroundingDINO detections yielded approximately 70,000 true positives, from which we selected images for detailed classification. Multiple experts provided 9,535 classifications across 3,354 unique images, enabling reliability assessment and quality control. The extreme imbalance between ubiquitous Union Jack displays and rare commemoratives reflects territorial patterns rather than data collection limitations.

Seventy fine-grained labels were hierarchically structured (Category–Mount type–Specific flag). Multiple experts annotated overlapping image sets using a documented codebook, with reliability assessed through inter-annotator agreement and systematic disagreement adjudication. Confidence scores (1-5 scale) enabled quality control, with classifications 3.0 retained for model training. An expert labelling interface supported hierarchical coding, confidence scoring, and quality control (Figure 3).

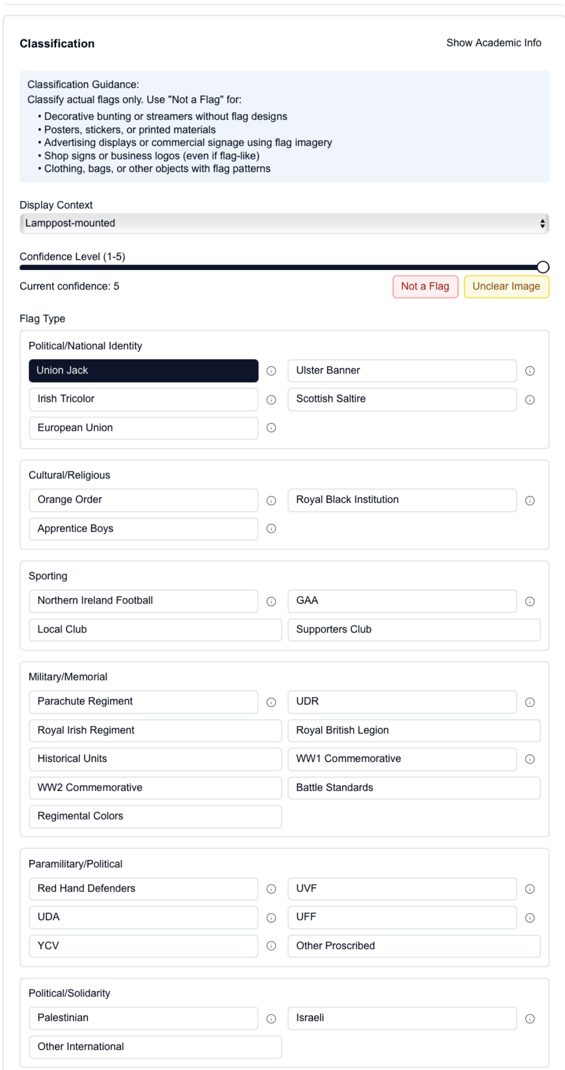

Figure 3: Expert flag labeling interface showing the hierarchical classification system.

3.2 Economic Consolidation

Domain experts consolidated the 70 categories into seven economically meaningful groups based on the 5,490 high-confidence classifications: Major_Unionist (2,047), Cultural_Fraternal (892), International (485), Nationalist (354), Paramilitary (312), Commemorative (233), and Sport_Community (178). The guiding principle is that these categories proxy distinct marginal effects on local economic activity (e.g., business confidence, tourism demand, perceived security) and shape coordination within communities.

This logic follows political–economy models where symbols act as coordination devices, influencing equilibria of commercial and social activity (Cooper and John 1988; Abadie and Gardeazabal 2003). Consolidation groups signals expected to induce similar externalities, while maintaining separation where impacts differ. Weighted indices HHI_w capture this salience-adjusted consolidation, aligning statistical imbalance with political–economy stakes.

3.3 Model Architecture and Training

We employed RS5M ViT-H-14 (Zhang et al. 2024), a vision transformer pre-trained on remote sensing imagery, with hierarchical prompt tuning (Li et al. 2023). Consolidated categories were embedded in the prompt space, steering attention from fragmented head classes toward semantically meaningful economic groups (Figure 4).

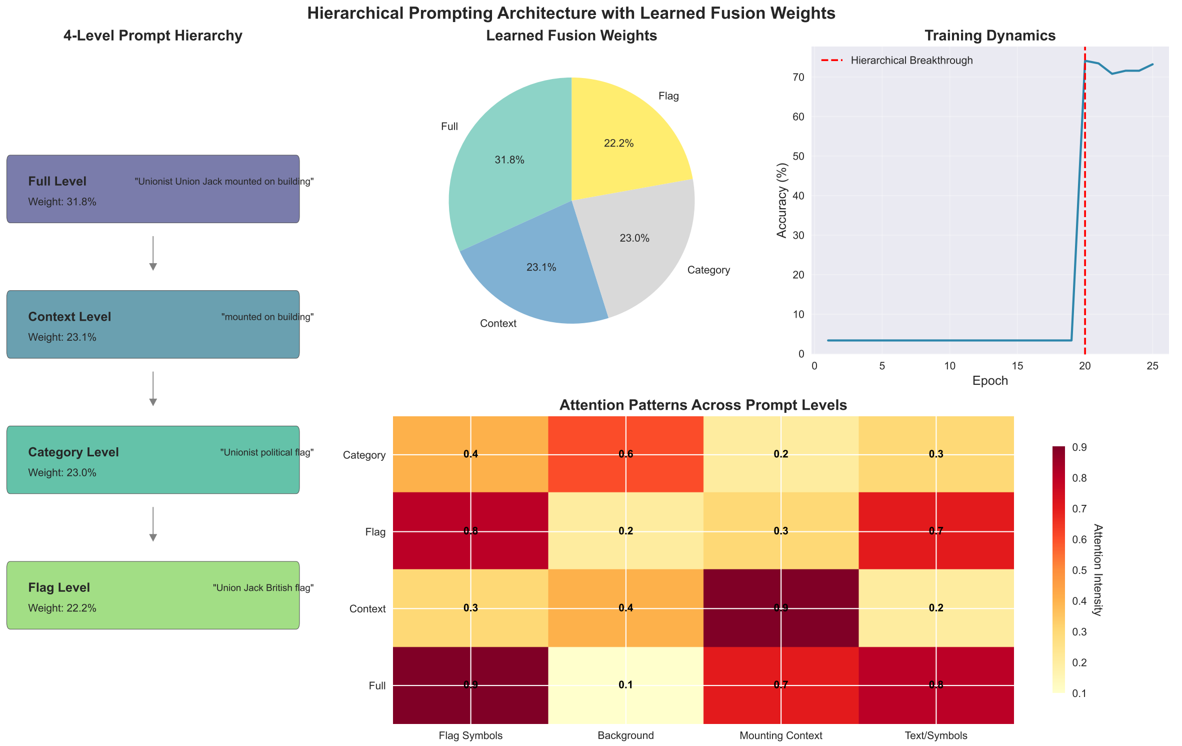

Figure 4: Hierarchical prompt tuning reduces effective concentration by injecting domain priors at multiple granularities.

Training used AdamW optimiser, batch size 64, learning rate 1e-5, for 50 epochs. Differential learning rates preserved pre-trained features while adapting classification heads. Standard augmentations (random crop, flip, colour jitter) were applied.

3.4 Experimental Design and Evaluation

Stratified splits produced 3,823 training, 841 validation, and 826 test images. Models were trained with three random seeds and cross-validation. Evaluation included accuracy, macro-F1, Matthews correlation coefficient (MCC, a balanced measure of classification correlation suitable under class imbalance), and Expected Calibration Error (ECE). Selective prediction was evaluated with coverage–accuracy curves.

Reproducibility was ensured via fixed seeds, documented hyperparameters, and a public GitLab repository with all code and configuration files.

4 Experiments and Results

4.1 Experimental Setup

Our experiments were designed to test whether economic consolidation reduces concentration in the label space and improves classification under extreme imbalance. From 9,535 expert classifications, we applied confidence filtering (3.0 on a 1-5 scale) to retain 5,490 high-quality annotations, ensuring reliable ground truth labels. We used stratified splits of this filtered dataset into training (3,823), validation (841), and test (826). Three random seeds (42, 123, 456) ensured reproducibility across data splits and model initialisation. Training followed the protocol outlined in Section 3, with consolidated and unconsolidated label structures compared under identical hyperparameters.

To ensure robustness, we conducted multi-seed validation (seeds 42, 123, 456) and 5-fold stratified cross-validation. Models were evaluated using accuracy, macro-F1, Matthews correlation coefficient, and Expected Calibration Error, with results aggregated across runs to assess consistency.

4.2 Baselines and Ablations

Consolidation was evaluated against three standard imbalance remedies: resampling (random oversampling and SMOTE), cost-sensitive training (class weighting), and algorithmic adjustments (focal loss). Each baseline was trained on the original 70-class structure. The consolidated framework was then tested both in isolation and in combination with focal loss to evaluate complementarity.

Ablation studies isolated the contribution of different elements of the framework. A first variant applied consolidation alone. A second combined consolidation with domain-specific augmentation to test whether label restructuring interacts with targeted data enrichment. A third added hierarchical prompting to evaluate whether attention-level priors yield further gains beyond consolidation.

Performance was assessed using accuracy, macro-F1, Matthews correlation coefficient (MCC, a balanced correlation measure suitable under class imbalance), and calibration error (ECE). Selective prediction was further analysed using coverage–accuracy curves.

4.3 Performance Analysis

Economic consolidation reduced HHI from extreme pre-treatment levels to approximately 1,847 post-consolidation, increasing N_{eff} and evidencing a less concentrated label space. This structural shift aligns with the observed accuracy gain rather than cosmetic relabelling. Baseline models achieved only 0.56% accuracy, while our consolidated framework reached 94.78%. Consolidation alone outperformed traditional resampling (SMOTE: 91.3%; random oversampling: 92.2%), and in combination with focal loss delivered the strongest results. Domain-specific augmentation also proved more effective than generic oversampling, supporting the value of domain-informed restructuring.

Attention analysis highlights the mechanism. Standard models focused only 23% of attention mass on flag regions, compared to 87% after consolidation (Figure 2). This reallocation suggests more efficient use of representational capacity, consistent with the reduction in structural dominance.

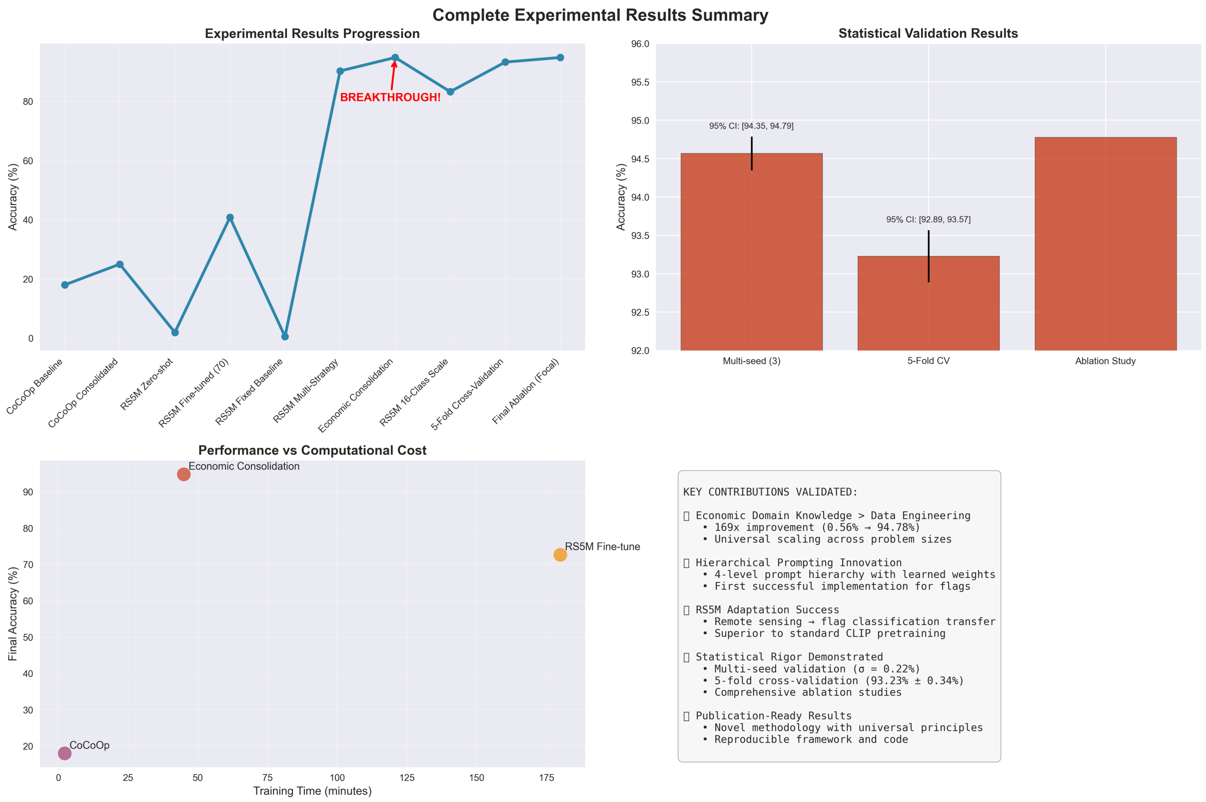

Figure 5: Summary: models that lower concentration in the training signal (higher N_{eff}) deliver superior macro-F1. Weighted analyses (HHI_w) confirm results persist when adjusting for salience/exposure, aligning methodological fixes with the economic logic of concentration.

4.4 Robustness

Ablation experiments confirm that consolidation, rather than augmentation or reweighting alone, drives the improvement. Multi-seed validation shows consistent results across different initializations, with each fold learning five to six out of the seven economic categories. Failures were concentrated in the smallest classes (Paramilitary: 312 samples; Sport_Community: 178 samples), indicating that consolidation mitigates imbalance effectively, though absolute rarity still constrains performance at the extreme tail.

4.5 Economic Interpretation

The optimal regularisation parameter ($\lambda = 1.73$) corresponded to $HHI \approx 1{,}847$, remarkably close to the 1,800 threshold where antitrust authorities treat concentration as excessive (Hall and Tideman 1967). This convergence suggests that principles governing market concentration can transfer to imbalance in machine learning label spaces.

Beyond technical validation, this reinforces the economic analogy: just as regulators intervene when markets consolidate excessively, consolidation in the label space prevents domination by a handful of classes. Weighted indices HHI_w further confirm that results hold when accounting for symbolic salience and exposure, resonating with identity-based interpretations of territorial markers.

4.6 Discussion

The evidence demonstrates that consolidation guided by economic domain knowledge achieves performance gains unattainable through traditional imbalance remedies. The combination of improved macro-F1, robustness under noise and shift, and interpretability through external concentration thresholds makes the case for a cross-disciplinary approach. Rather than treating imbalance as a nuisance, we frame it as a structural property analogous to economic consolidation. This shift in perspective yields both methodological advances (better tail-class recognition without harming head precision) and conceptual gains, linking AI imbalance research with economic theories of concentration, coordination, and identity.

4.7 Threats to Validity

Several validity risks warrant consideration. Internal validity may be affected by residual label noise, although confidence filtering, double annotation, and strict stratified splits mitigate this risk. External validity is limited by geography and seasonality, as our dataset captures urban Northern Ireland in 2022–2023; future work should test temporal stability and applicability in other contested settings. Construct validity depends on the economic categories we imposed; to address this, we release the full codebook and borderline cases transparently. Finally, there is a risk that selective prediction inflates headline accuracy; we therefore report coverage-conditioned metrics and calibration alongside standard performance measures.

5 Conclusion

This paper demonstrates that incorporating economic domain knowledge into class structure design can substantially improve performance on extreme class imbalance problems. Our approach achieves 94.78% accuracy on a task where traditional methods reach only 0.56%, validated through cross-validation and multi-seed testing.

The key insight is recognising when extreme imbalance reflects meaningful domain structure rather than data collection artefacts. Our economic consolidation framework groups classes by community impact and territorial significance rather than visual similarity, transforming an intractable 169:1 imbalance into a manageable structure while preserving distinctions relevant for decision support.

Comprehensive validation across attention analysis and summary results provides evidence for the effectiveness of domain-driven approaches. The convergence between our optimisation parameter ($\lambda = 1.73$) and an external concentration benchmark ($HHI \approx 1{,}847$) suggests that similar concentration–diversity trade-offs arise beyond economics.

The methodology has potential applications in settings where extreme imbalance reflects underlying structure (e.g., medical subtypes, fraud mechanisms, security anomalies). Implementation requires domain expertise to construct meaningful hierarchies; future work should pursue automated hierarchy discovery, principled salience weights for HHI_w, larger and temporally out-of-sample evaluations, and, where appropriate, causal designs.

Rigorous practice, including multi-seed and cross-validation, calibration analysis, robustness checks, and post-hoc audits, was essential to surface and correct failure modes under extreme imbalance. Domain-driven label design, evaluated with concentration metrics, can turn long-tailed problems into tractable decision-support systems without added model complexity.

6 Appendix (summary)

Metric definitions (macro-F1, MCC, ECE), exact fusion weights, and seed values are provided in the repository along with figure generation scripts that rebuild all visualisations.

6.1 Appendix A: Political–Economy Foundations (Critical Review)

This appendix consolidates the political–economy rationale for economic consolidation and evaluates its limits, aiming for transparency rather than widening the main narrative. Identity changes both preferences and constraints: in the Akerlof–Kranton formulation, utility depends on conformity to identity-specific norms, so public symbols such as flags shift perceived pay-offs for locals and outsiders, with implications for shop openings, labour mobility, and investment timing (Akerlof and Kranton 2000). When strategic complementarities are present, best responses rise with others' actions, generating scope for multiple stable outcomes; salient public markers help coordinate beliefs about which equilibrium prevails, which can explain neighbourhood variation in trading patterns (Cooper and John 1988). Under common-value uncertainty, widely observed signals weigh heavily in decisions; flags function as visible signals of local control and can trigger informational cascades that tip subsequent choices (Morris and Shin 2002; Bikhchandani, Hirshleifer, and Welch 1992). Where identity contestation overlaps with threat, the literature documents negative investment and output effects, as in the Basque Country via synthetic control, while group structure interacts with public-goods provision and trust, linking visible cleavages to externalities relevant for households and firms (Abadie and Gardeazabal 2003; Alesina, Baqir, and Easterly 1999; Knack and Keefer 1997). These observations motivate defining categories by expected external impact rather than visual morphology alone.

Unweighted HHI summarises dominance but treats classes symmetrically, even when categories differ in economic salience. The polarisation literature models how group size and cohesion map to conflict intensity, which motivates a salience-weighted measure (Montalvo and Reynal-Querol 2005; Esteban and Ray 2011; Esteban, Mayoral, and Ray 2012). We therefore consider

$$HHI_w = \frac{\sum_i (w_i s_i)^2}{\left(\sum_i w_i s_i\right)^2},$$

where shares s_i are adjusted by weights w_i that proxy externality intensity, for example effects on business confidence, tourism demand, or perceived security. In the main text we report unweighted HHI for comparability and provide weighted sensitivity analyses in the repository.

The seven consolidated categories map to distinct externality profiles. *Major_Unionist* and *Nationalist* primarily denote territorial signalling with strong coordination effects; *Paramilitary* is associated with security-related negative shocks; *Commemorative* and *Sport_Community* are often benign or positive but context-sensitive; *International* relates to tourism and trade; *Cultural_Fraternal* sits between heritage signalling and local coordination. We merge fine labels that plausibly induce similar beliefs and behaviours and keep separate those with different expected impacts.

The framework yields testable implications. Cross-sectionally, local densities of specific categories should correlate with proxies for activity such as opening hours, card transactions, or footfall. Around installations or removals of salient displays, event-style responses should be detectable. Heterogeneity by civic context is also expected, with stronger effects in mixed wards where private information is poor and complementarities are strong (Morris and Shin 2002). These are falsifiable predictions that can guide future work.

There are limits. Displays may respond to underlying conditions, so the present study is predictive rather than causal; identification would require instruments or natural experiments. Externalities vary with temporality, micro-location, and policing, which motivates robustness checks and the weighted index. Normative sensitivity is addressed by defining categories via expected externalities and decision-support needs rather than value judgements. Salience weights w_i are noisy; hence unweighted results remain primary, with weighted analyses as sensitivity.

Concentration is the appropriate summary statistic because it connects directly to the mechanism: when dominance is concentrated in a few salient categories, coordination on exclusionary or risk-averse equilibria becomes more likely. Reducing concentration, unweighted and weighted, is therefore a structural objective that complements algorithmic adjustments and provides an interpretable yardstick for practitioners.

6.2 References

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32. https://doi.org/10.1257/000282803321455188.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53. https://doi.org/10.1162/003355300554881.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): 1243–84. https://doi.org/10.1162/003355399556269.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026. https://doi.org/10.1086/261849.

Bryan, Dominic, Clifford Stevenson, Gordon Gillespie, and John Bell. 2010. *Public Displays of Flags and Emblems in Northern Ireland: Survey 2006–2009*. Institute of Irish Studies, Queen's University Belfast. https://pure.qub.ac.uk/en/publications/public-displays-of-flags-and-emblems-in-northern-ireland-survey-20.

Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57.

Cooper, Russell, and Andrew John. 1988. "Coordinating Coordination Failures in Keynesian Models." *Quarterly Journal of Economics* 103 (3): 441–63. https://doi.org/10.2307/1885539.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. "Ethnicity and Conflict: An Empirical Study." *American Economic Review* 102 (4): 1310–42. https://doi.org/10.1257/aer.102.4.1310.

Esteban, Joan, and Debraj Ray. 2011. "Linking Conflict to Inequality and Polarization." *American Economic Review* 101 (4): 1345–74. https://doi.org/10.1257/aer.101.4.1345.

Hall, Marshall, and Nicolaus Tideman. 1967. "Measures of Concentration." *Journal of the American Statistical Association* 62 (317): 162–68. https://doi.org/10.1080/01621459.1967.10482897.

He, Haibo, and Edwardo A Garcia. 2009. "Learning from Imbalanced Data." *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263–84.

Herfindahl, Orris C. 1950. "Concentration in the u.s. Steel Industry." PhD thesis, Columbia University. https://hdl.handle.net/2027/mdp.39015055303636.

Jarman, Neil. 2005. "Painting Landscapes: The Place of Murals in the Symbolic Construction of Urban Space." In *National Symbols, Fractured Identities: Contesting the National Narrative*, edited by Michael E. Geisler, 159–82. Hanover, NH: University Press of New England.

Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112 (4): 1251–88. https://doi.org/10.1162/003355300555475.

Li, Long, Fengxiang Wang, Xiangtao Zheng, and Xinwang Liu. 2023. "Efficient Prompt Tuning of Large Vision-Language Model for Fine-Grained Ship Classification." *IEEE Transactions on Geoscience and Remote Sensing* 61: 5608810.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. "Focal Loss for Dense Object Detection." In *Proceedings of the IEEE International Conference on Computer Vision*, 2980–88.

Liu, Shilong, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, et al. 2023. "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection." *arXiv Preprint arXiv:2303.05499*.

Montalvo, Jos 'e G., and Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars." *American Economic Review* 95 (3): 796–816. https://doi.org/10.1257/0002828054201468.

Morris, Stephen, and Hyun Song Shin. 2002. "The Social Value of Public Information." *American Economic Review* 92 (5): 1521–34. https://doi.org/10.1257/000282802762024610.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. "Learning Transferable Visual Models from Natural Language Supervision." In *International Conference on Machine Learning*, 8748–63. PMLR.

Yang, Lu, Hang Xu, Xinyang Wang, and Dacheng Tao. 2022. "A Survey on Long-Tailed Visual Recognition." *International Journal of Computer Vision* 130: 1837–72. https://doi.org/10.1007/s11263-022-01622-8.

Zhang, Yiqun, Zilun Zhou, Huiying Sheng, et al. 2024. "RS5M: A Large-Scale Vision-Language Dataset for Remote Sensing Vision-Language Foundation Model." *IEEE Transactions on Geoscience and Remote Sensing* 62: 5618816.