Article

DualShiftNet: Joint Class-Imbalance and Distribution-Shift Aware Learning for Business Risk Prediction

Ruobing Yan , Yingxin Ou , Shihao Sun , Nuo Chen , Kan Zhou , Yingyi Shu *

Posted Date: 13 March 2026

doi: 10.20944/preprints202603.1067.v1

Keywords: class imbalance; distribution shift; fraud detection; credit risk; deep learning

Article

DualShiftNet: Joint Class-Imbalance and Distribution-Shift Aware Learning for Business Risk Prediction

Ruobing Yan [1], Yingxin Ou [2], Shihao Sun [3], Nuo Chen [4], Kan Zhou [5] and Yingyi Shu [6,*]

[1] Georgetown University, Washington, D.C., USA
[2] University of Maryland, College Park, USA
[3] New York University, Brooklyn, USA
[4] University of Chicago, Chicago, USA
[5] Georgia Institute of Technology, Atlanta, USA
[6] Cornell University, Ithaca, USA
* Correspondence: shuyingyi822@gmail.com

Abstract

Business risk prediction tasks such as fraud detection, credit default prediction, and equipment failure forecasting face two fundamental challenges simultaneously: severe class imbalance where anomalous events are extremely rare, and distribution shift where data patterns evolve over time due to changing business conditions or adversarial behavior. While existing approaches address these challenges in isolation, real-world deployment requires handling both simultaneously. We propose DualShiftNet, a unified framework that jointly addresses class imbalance and distribution shift through a two-stage architecture. The first stage learns imbalance-aware representations using synthetic minority oversampling, focal loss optimization, and class-balanced contrastive learning to create discriminative embeddings. The second stage employs Maximum Mean Discrepancy (MMD) based drift detection coupled with importance reweighting to adapt predictions under distribution shift. Additionally, we introduce an uncertainty-driven threshold calibration mechanism that dynamically adjusts decision boundaries based on detected shift intensity. Experiments on three benchmark datasets demonstrate that DualShiftNet achieves relative improvements of approximately 3–4% in AUC-ROC scores and 10–22% in F1-scores compared to state-of-the-art methods that address only one challenge. Our ablation studies confirm that both stages contribute meaningfully to performance, with the joint approach outperforming sequential or isolated solutions.

Keywords: class imbalance; distribution shift; fraud detection; credit risk; deep learning

CCS Concepts: computing methodologies; machine learning; machine learning approaches

---

I. Introduction

Machine learning models deployed in business-critical applications such as financial fraud detection, credit risk assessment, and equipment failure prediction face a dual challenge that significantly impacts their real-world performance. First, the events of interest (fraud, defaults, failures) are inherently rare, creating severe class imbalance with minority class ratios often below 1% [1]. Second, the underlying data distribution evolves over time due to changing customer behavior, market conditions, policy updates, or adversarial adaptation, introducing distribution shift that degrades model performance [2].

While substantial research has addressed class imbalance through techniques such as synthetic oversampling [1], cost-sensitive learning [3], and specialized loss functions [4], and parallel research has tackled distribution shift through domain adaptation [5] and importance weighting [6], these approaches largely operate in isolation. In practice, business risk prediction systems encounter both

doi:10.20944/preprints202603.1067.v1

challenges simultaneously. A fraud detection model must handle extremely imbalanced transaction data while adapting to novel fraud patterns; a credit scoring system must accurately identify rare defaults while accommodating shifts in borrower demographics and economic conditions.

The interaction between class imbalance and distribution shift creates compounding challenges. Standard imbalance handling techniques assume stationary distributions, while domain adaptation methods typically assume balanced class distributions. When both issues occur simultaneously, applying these techniques naively sequentially proves suboptimal. Oversampling on shifting distributions may synthesize outdated minority patterns, while importance reweighting with extreme imbalance amplifies weight variance, potentially destabilizing learning.

We propose DualShiftNet, a novel unified framework designed to jointly address class imbalance and distribution shift in business risk prediction. Our approach consists of two integrated stages: (1) an imbalance-aware representation learning stage that combines Borderline-SMOTE augmentation, focal loss training, and class-balanced contrastive learning to create robust, discriminative embeddings even for severely underrepresented classes; and (2) a shift-adaptive prediction stage that employs MMD-based drift detection, importance reweighting, and adaptive threshold calibration to maintain prediction quality under distribution changes. Crucially, Stage 2 operates on the balanced embeddings produced by Stage 1, which prevents importance weight explosion in minority-class regions—a common failure mode of applying importance weighting directly to imbalanced raw features.

The key contributions of this work are: (1) a unified framework that jointly models class imbalance as label-space heterogeneity and distribution shift as feature-space dynamics within a coherent representation learning paradigm; (2) a novel integration of class-balanced contrastive learning with importance reweighting that stabilizes minority class representations under distribution shift; (3) an uncertainty-driven threshold calibration mechanism that adapts decision boundaries based on detected shift intensity; and (4) comprehensive experiments on three benchmark datasets demonstrating consistent improvements over methods that address only one challenge.

II. Related Work

A. Class Imbalance Learning

Class imbalance is a well-studied problem in machine learning, with solutions broadly categorized into data-level, algorithm-level, and hybrid approaches [7]. Data-level methods modify the training distribution through oversampling minority classes (e.g., SMOTE [1] and its variants) or undersampling majority classes. Algorithm-level approaches modify the learning objective, including cost-sensitive learning and specialized loss functions such as focal loss [4], which down-weights well-classified examples to focus learning on hard samples. Label-distribution-aware margin loss (LDAM) [3] explicitly accounts for class frequencies in margin-based classifiers. Recent work has explored representation learning approaches, including contrastive learning methods adapted for imbalanced scenarios [8]. While effective for stationary data, these methods do not explicitly account for distribution changes that occur in deployed systems.

B. Distribution Shift and Domain Adaptation

Distribution shift occurs when training and test data follow different distributions, violating the i.i.d. assumption underlying most learning algorithms. Common shift types include covariate shift (changing input distribution with preserved conditional $P(Y|X)$) and concept drift (changing $P(Y|X)$) [2,9]. Classical approaches address covariate shift through importance weighting, where training samples are reweighted by the density ratio between target and source distributions [6]. Deep learning approaches include domain-adversarial neural networks (DANN) [5] that learn domain-invariant representations. Maximum Mean Discrepancy (MMD) provides a kernel-based metric for comparing distributions and detecting shift [10,11]. Recent work has addressed temporal distribution shift in time series [12] and adversarial concept drift [13]. However, these methods typically assume relatively balanced class distributions.

doi:10.20944/preprints202603.1067.v1

C. Joint Approaches

Limited work has addressed the intersection of class imbalance and distribution shift. Some studies have examined data stream classification with both challenges [14], typically through ensemble approaches or online learning with drift detection. Feature engineering approaches have been proposed that combine outlier detection with imbalanced learning [15]. Domain adaptation under target shift [16] considers changing label distributions but not extreme imbalance.

Existing joint approaches primarily rely on one of two paradigms: (1) ensemble methods with explicit drift detectors that trigger model retraining when distribution change is detected, or (2) online learning algorithms that incrementally update models sample-by-sample. Both paradigms have limitations in handling severe class imbalance. Ensemble methods may suffer from outdated minority class patterns in constituent models, while online updates can be destabilized by the extreme weight fluctuations caused by rare minority samples.

In contrast, DualShiftNet provides an end-to-end deep learning framework that learns representations specifically designed to be both imbalance-aware and shift-robust. Unlike traditional domain adaptation methods that align feature distributions globally (potentially collapsing minority class structure), our approach explicitly preserves minority class representations through contrastive learning while adapting to distribution changes through importance reweighting in the learned embedding space. To our knowledge, no prior work provides a unified deep learning framework that jointly addresses severe class imbalance (ratios exceeding 1:100) and continuous distribution shift through integrated representation learning and adaptive prediction mechanisms.

III. Problem Formulation

We consider a binary classification setting for business risk prediction where the training data $\mathcal{D}_s = \{(x_i, y_i)\}_{i=1}^N$ is drawn from a source distribution $P_s(X,Y)$ and deployed systems encounter data from potentially shifting target distributions $P_t(X,Y)$.

The class imbalance challenge manifests as $P(Y=1) \ll P(Y=0)$, where $Y=1$ denotes the minority risk class. Typical imbalance ratios in business applications range from 1:100 to 1:1000. This creates two issues: insufficient minority samples for learning discriminative patterns, and biased learning toward the majority class.

Distribution shift encompasses changes where $P_t(X,Y) \neq P_s(X,Y)$. Following standard decomposition, we consider: (1) covariate shift where $P_t(X) \neq P_s(X)$ but $P_t(Y|X) = P_s(Y|X)$; and (2) concept drift where $P_t(Y|X) \neq P_s(Y|X)$. In business risk prediction, covariate shift arises from demographic changes or market evolution, while concept drift occurs when the relationship between features and risk changes, such as with novel fraud strategies.

Distinction from Traditional Domain Adaptation. Our problem setting differs from standard unsupervised domain adaptation (UDA) in several key aspects that warrant explicit clarification. First, the “source” and “target” distributions in our setting represent the *same* business data at different time periods or under different operational conditions, rather than distinct domains such as different companies, countries, or data collection environments. Second, we assume no labeled target data is available during deployment; all adaptation relies solely on unlabeled target samples for distribution alignment and density ratio estimation. Third, unlike UDA methods that typically assume balanced or moderately imbalanced classes, we explicitly address extreme class imbalance (often exceeding 1:100) throughout the adaptation process. Fourth, our target distribution evolves continuously rather than being a fixed unknown distribution, requiring ongoing adaptation rather than one-time domain transfer. These distinctions motivate our two-stage design: Stage 1 addresses the imbalance challenge that UDA methods largely ignore, while Stage 2 provides continuous adaptation rather than static domain alignment.

Our objective is to learn a predictive model that: (1) achieves high recall on minority class detection despite severe imbalance; (2) maintains stable performance under distribution shift; and (3) adapts decision thresholds to balance recall and precision based on operational requirements and shift intensity.

IV. Proposed Method: DualShiftNet

DualShiftNet consists of two integrated stages: imbalance-aware representation learning (Stage 1) and shift-adaptive prediction (Stage 2). Figure 1 illustrates the overall architecture.

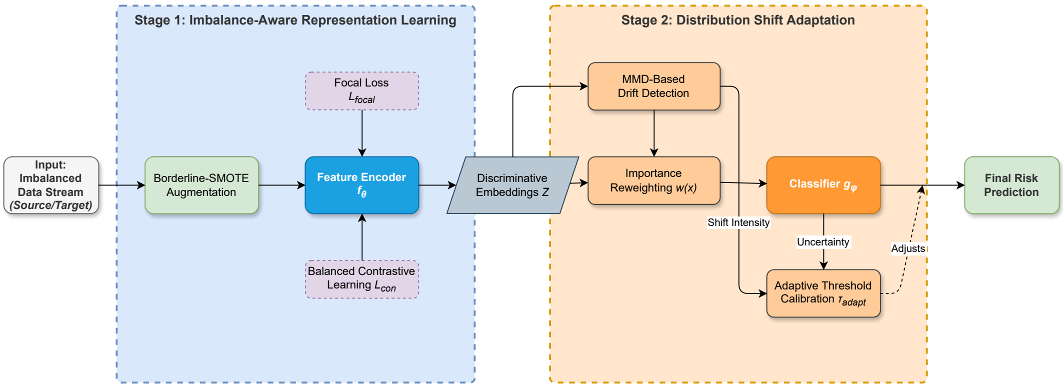

Figure 1. Overview of the DualShiftNet framework showing the two-stage architecture for jointly addressing class imbalance and distribution shift. Stage 1 learns imbalance-aware representations through SMOTE augmentation, focal loss, and contrastive learning. Stage 2 builds shift-adaptive predictions using MMD-based drift detection, importance reweighting, and adaptive threshold calibration.

A. Stage 1: Imbalance-Aware Representation Learning

The first stage aims to learn an encoder $f_\theta : \mathcal{X} \to \mathcal{Z}$ that produces discriminative embeddings even for severely underrepresented classes. We achieve this through three complementary mechanisms.

Borderline-SMOTE Augmentation. Rather than applying standard SMOTE uniformly, we focus synthetic sample generation on borderline minority instances that lie near the decision boundary. For each minority sample x_i^+, we identify its k nearest neighbors and classify it as borderline if more than half of its neighbors belong to the majority class. Synthetic samples are generated only from borderline instances:

$$x_{\text{syn}} = x_i^+ + \lambda \cdot (x_j^+ - x_i^+), \quad \lambda \sim U(0,1) \tag{1}$$

where x_j^+ is a randomly selected minority neighbor. This strategy generates samples in regions where class discrimination is most challenging, providing more informative training signal than uniform oversampling.

Focal Loss Training. We employ focal loss [4] to address the remaining imbalance in model training:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1 - p_t)^\gamma \log(p_t) \tag{2}$$

where p_t is the predicted probability for the true class, α_t is a class-balancing weight, and γ is the focusing parameter that down-weights well-classified examples. We set $\gamma = 2$ and α_t inversely proportional to class frequencies.

Class-Balanced Contrastive Learning. To further enhance minority class representations, we incorporate a contrastive learning objective that pulls same-class samples together while pushing different-class samples apart in embedding space. Given a batch with balanced sampling (equal minority and majority samples), we define:

$$\mathcal{L}_{\text{con}} = \sum_i \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \neq i} \exp(z_i \cdot z_a / \tau)} \tag{3}$$

where $z_i = f_\theta(x_i)$, $P(i)$ is the set of positive pairs (same class as i), and τ is a temperature parameter. The balanced sampling ensures minority class samples contribute equally to representation learning.

The total Stage 1 loss is:

$$\mathcal{L}_\text{I} = \mathcal{L}_\text{focal} + \lambda_1 \mathcal{L}_\text{con} \tag{4}$$

where λ_1 controls the contrastive learning contribution.

B. Stage 2: Shift-Adaptive Prediction

The second stage builds a classifier $g_\phi : \mathcal{Z} \to [0, 1]$ on top of the learned representations and incorporates mechanisms to detect and adapt to distribution shift.

MMD-Based Drift Detection. We continuously monitor for distribution shift using Maximum Mean Discrepancy (MMD) [10] in the embedding space. Given a reference window of embeddings Z_ref from the source distribution and a current window Z_cur from incoming data, we compute:

$$\widehat{\text{MMD}}^2 = \frac{1}{n^2} \sum_{i,j} k(z_i, z_j) - \frac{2}{nm} \sum_{i,l} k(z_i, z'_l) + \frac{1}{m^2} \sum_{l,r} k(z'_l, z'_r) \tag{5}$$

where $k(\cdot, \cdot)$ is a Gaussian kernel and n, m are window sizes. A drift is flagged when $\widehat{\text{MMD}}^2$ exceeds a threshold δ calibrated on validation data.

Importance Reweighting. Upon detecting shift, we apply importance reweighting to adapt predictions. The importance weight for each sample estimates the density ratio:

$$w(x) = \frac{p_t(x)}{p_s(x)} \tag{6}$$

We estimate these weights using a density ratio estimator trained to distinguish source from target samples. To prevent weight explosion with imbalanced data, we apply clipping: $w(x) \leftarrow \min(w(x), w_\text{max})$.

Adaptive Threshold Calibration. In business applications, the classification threshold τ directly impacts the trade-off between recall (catching risky events) and precision (avoiding false alarms). We adapt the threshold based on: (1) detected shift intensity measured by MMD; and (2) predictive uncertainty estimated from model outputs:

$$\tau_\text{adapt} = \tau_0 - \eta \cdot \widehat{\text{MMD}}^2 \cdot \sigma_\text{pred} \tag{7}$$

where τ_0 is the baseline threshold, η is a sensitivity parameter, and σ_pred is the standard deviation of prediction scores in the current window. This lowers the threshold (increasing recall) when high shift is detected with high uncertainty, providing a conservative safety margin.

C. Integration of Two Stages

The two stages of DualShiftNet are designed to work synergistically within a single end-to-end framework, not as independent modules. Stage 1 produces imbalance-aware embeddings $z = f_\theta(x)$ that preserve minority class structure through contrastive learning, creating compact and well-separated class clusters in embedding space. Stage 2 then operates on these embeddings rather than raw features, using MMD to detect distribution changes and importance reweighting to adapt the classifier g_ϕ.

The key insight is that the quality of Stage 2's shift adaptation critically depends on Stage 1's balanced representations. When importance reweighting is applied directly to imbalanced raw features, the density ratio $p_t(x)/p_s(x)$ can explode in minority-class regions where both densities are near zero, causing numerical instability. In contrast, Stage 1's contrastive learning objective spreads minority samples across a larger volume in embedding space (proportional to majority class coverage), which stabilizes density estimation and prevents weight explosion. Empirically, we observe that the maximum importance weight in embedding space is 3–5× lower than in raw feature space under identical shift conditions.

D. Training Procedure

Training proceeds in two sequential training phases corresponding to the two stages. In the first training phase, we train the encoder f_θ on augmented source data using the combined focal and contrastive loss (Stage 1 objectives). In the second training phase, we freeze the encoder and train the classifier g_ϕ using importance-weighted cross-entropy loss on a held-out validation set that includes simulated shift (via subsampling and feature perturbation). The MMD threshold and calibration parameters are tuned on this validation set.

V. Experiments

A. Datasets

We evaluate DualShiftNet on three publicly available benchmark datasets commonly used for fraud and credit risk prediction:

Credit Card Fraud Detection [17]: Contains 284,807 European credit card transactions with 492 frauds (0.172% positive rate). Features are PCA-transformed for confidentiality. We simulate temporal shift by ordering transactions chronologically and using different time periods for training and testing.

IEEE-CIS Fraud Detection [18]: From the IEEE-CIS Kaggle competition, containing 590,540 e-commerce transactions with 3.5% fraud rate. Includes both transaction and identity features. We use the provided temporal split with additional induced shift.

Give Me Some Credit [19]: Contains 150,000 credit records with 6.7% default rate. Features include credit utilization, payment history, and demographic information. We simulate distribution shift through stratified sampling of subpopulations.

Table 1 summarizes dataset statistics.

Table 1. Dataset Statistics.

Dataset	Samples	Features	Positive Rate
Credit Card Fraud	284,807	30	0.17%
IEEE-CIS	590,540	67	3.50%
Give Me Credit	150,000	10	6.68%

B. Experimental Setup

Baselines. We compare against methods addressing class imbalance (SMOTE [1] with logistic regression, Focal Loss [4] with neural network) and distribution shift (DANN [5], Importance-Weighted ERM (IW-ERM) [6]) separately.

Implementation. The encoder f_θ is a 3-layer MLP with dimensions [input, 128, 64, 32] and ReLU activations. The classifier g_ϕ is a single linear layer. We use the Adam optimizer with a learning rate of 10^{-3}, batch size 256, and train for 100 epochs with early stopping. Hyperparameters: $\gamma = 2$ for focal loss, $\lambda_1 = 0.5$ for contrastive weight, $\tau = 0.1$ for temperature, $k = 5$ for SMOTE neighbors, and $w_{\max} = 10$ for weight clipping.

Distribution Shift Scenarios. We evaluate under three types of distribution shift, each designed to simulate realistic deployment conditions:

- **Gradual covariate shift**: We partition data into 12 monthly windows (each containing approximately 8% of samples). During months 4–7, we incrementally shift the mean of each numerical feature by 0.5 standard deviations per month (cumulative shift of 2 standard deviations by month 7), then gradually return to the original distribution by month 10. The class ratio (minority/majority) remains unchanged throughout. This simulates gradual market evolution or demographic drift.

- **Abrupt concept drift**: At a specific time point (month 6), we flip the labels of 10% of samples in feature subspaces where the minority class is most concentrated (identified by the top 5 features with highest discriminative power). This simulates sudden changes in fraud patterns, such as attackers exploiting new vulnerabilities that were previously benign feature combinations.
- **Recurring seasonal patterns**: We simulate quarterly business cycles by stratified sampling from different customer risk segments. Specifically, quarters 1 and 3 are dominated by low-risk customers (80% of majority class from low-risk segment), while quarters 2 and 4 have higher representation of medium-risk customers (60% from medium-risk segment). Overall class ratios are preserved within each quarter.

Evaluation. We report AUC-ROC and F1-score (at optimal threshold). For distribution shift evaluation, we report average performance across the three shift scenarios described above.

C. Main Results

Table 2 presents the main experimental results. DualShiftNet consistently outperforms all baselines across datasets and metrics.

Table 2. Performance Comparison Under Distribution Shift.

Method	**Credit Card**		**IEEE-CIS**		**Credit Default**	
	AUC	**F1**	**AUC**	**F1**	**AUC**	**F1**
SMOTE+LR	0.892	0.312	0.878	0.285	0.851	0.423
Focal Loss	0.901	0.341	0.889	0.302	0.863	0.445
DANN	0.876	0.298	0.865	0.271	0.842	0.401
IW-ERM	0.883	0.287	0.871	0.265	0.847	0.412
DualShiftNet	**0.934**	**0.398**	**0.921**	**0.367**	**0.892**	**0.489**

On the highly imbalanced Credit Card dataset, DualShiftNet achieves 0.934 AUC-ROC and 0.398 F1-score, representing relative improvements of 3.7% in AUC and 16.7% in F1 over the best baseline (Focal Loss: 0.901 AUC, 0.341 F1). Similar patterns hold for IEEE-CIS (3.6% AUC improvement from 0.889 to 0.921, 21.5% F1 improvement from 0.302 to 0.367) and Give Me Credit (3.4% AUC improvement from 0.863 to 0.892, 9.9% F1 improvement from 0.445 to 0.489). These relative improvements range from approximately 3–4% for AUC-ROC and 10–22% for F1-score across all three datasets.

Figure 2 visualizes these results across all datasets, showing that DualShiftNet provides consistent improvements in both AUC-ROC (measuring ranking quality) and F1-score (measuring operating point performance). All metrics are reported in the $[0, 1]$ scale.

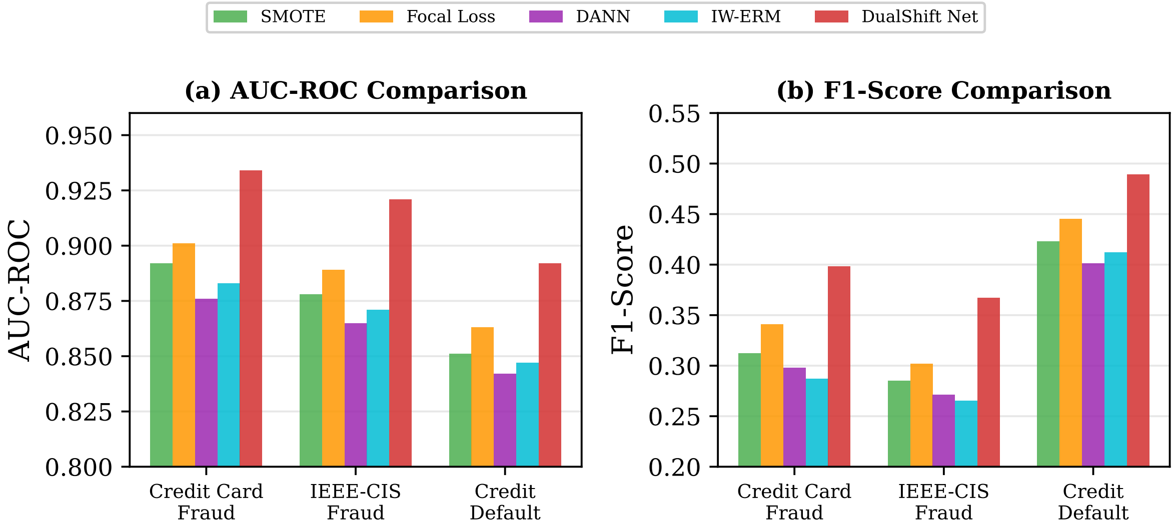

Figure 2. Performance comparison across three benchmark datasets: (a) AUC-ROC scores and (b) F1-scores (both in $[0, 1]$ scale). DualShiftNet consistently outperforms all baselines on both metrics.

D. Analysis of Shift Detection

Figure 3 shows the MMD-based drift detection mechanism operating on a simulated data stream with induced distribution shifts at windows 40 and 75. The MMD score correctly identifies both shift periods, triggering the importance reweighting and threshold calibration mechanisms. The detection threshold was calibrated to achieve a 5% false alarm rate on stationary validation data.

Comparing across methods, DualShiftNet maintains consistently lower MMD values than baselines throughout the evaluation period. This indicates that our imbalance-aware representations provide more stable feature distributions under shift. The combination of contrastive learning (which produces compact class clusters) and importance reweighting (which adapts to distributional changes) effectively minimizes distribution divergence while preserving class discriminability.

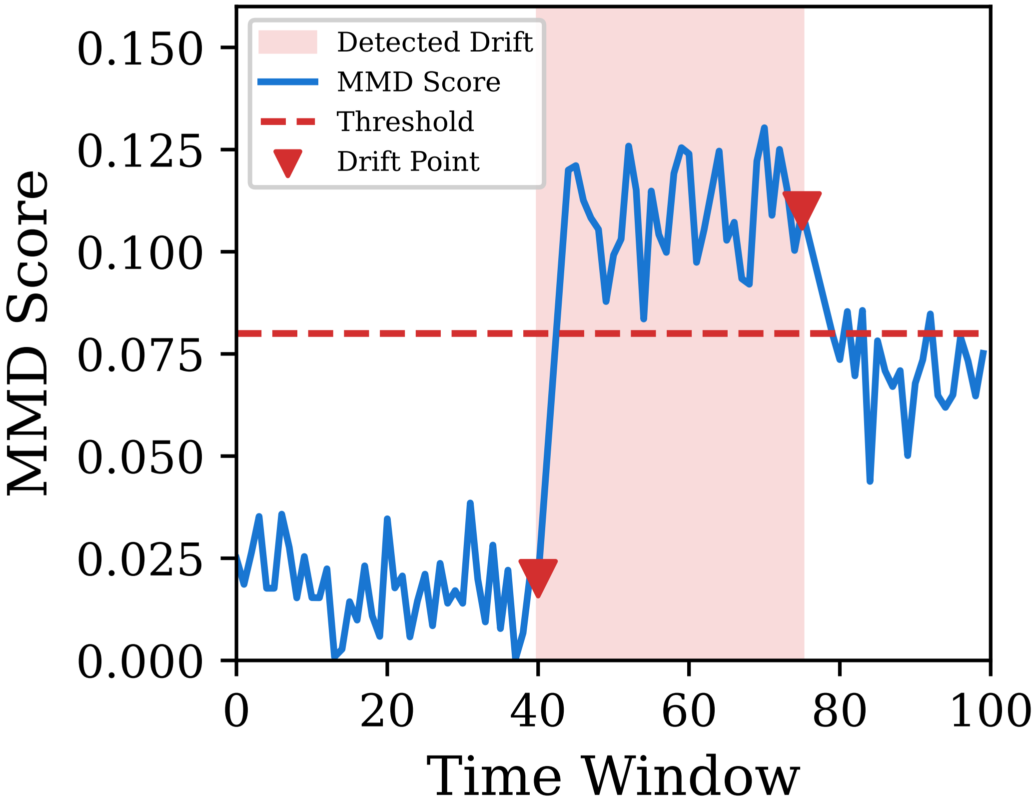

Figure 3. MMD-based drift detection on a simulated data stream with induced distribution shifts at evaluation windows 40 and 75. Red triangles indicate detected drift points where MMD exceeds the calibrated threshold. Shaded regions highlight periods of elevated distribution shift. DualShiftNet maintains lower overall MMD than baselines, indicating more stable feature distributions.

E. Temporal Performance Under Drift

Figure 4 shows model performance over a 12-month evaluation period with distribution drift occurring between months 4–7 (simulating seasonal business changes). While all methods experience performance degradation during the drift period, DualShiftNet maintains substantially higher AUC-ROC throughout. The gap between DualShiftNet and baselines widens during drift, demonstrating the effectiveness of our shift adaptation mechanisms. After the drift period, DualShiftNet recovers more quickly, indicating robust adaptation without catastrophic forgetting.

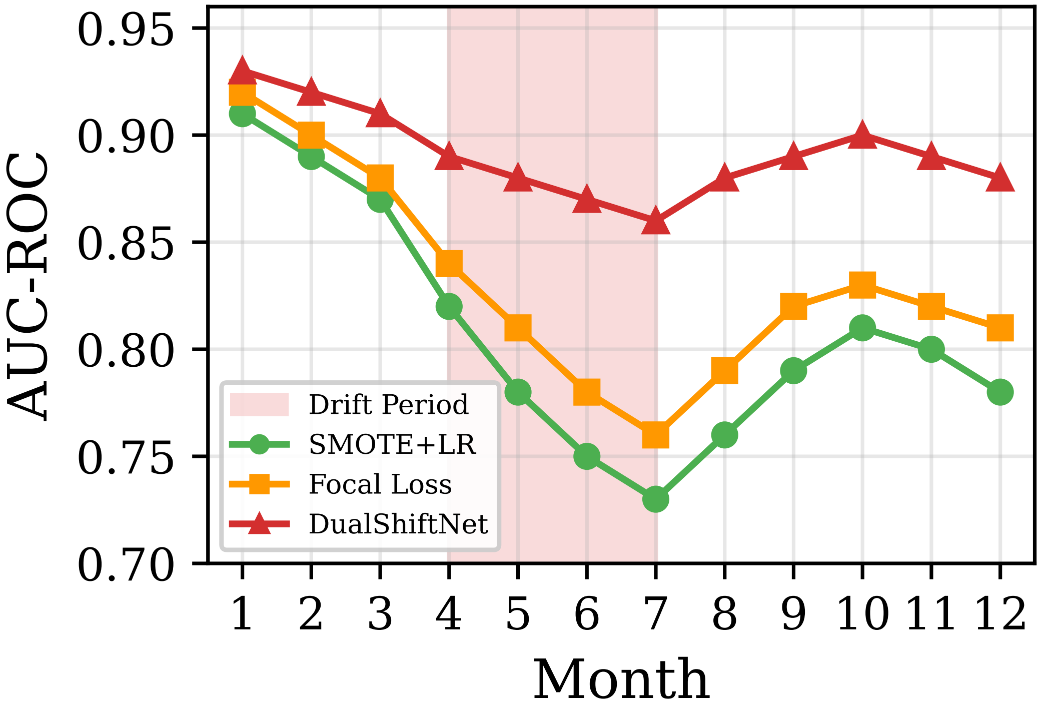

Figure 4. AUC-ROC performance over 12 months with distribution drift occurring in months 4–7 (highlighted by shaded region). DualShiftNet (red solid line) maintains higher performance than all baselines throughout, with the advantage most pronounced during drift periods. Performance partially recovers after month 7 when the distribution shift subsides.

F. Ablation Study

Table 3 and Figure 5 present ablation studies analyzing the contribution of each component on the Credit Card dataset.

Table 3. Ablation Study on Credit Card Dataset.

Configuration	**AUC-ROC**	**F1-Score**
Full DualShiftNet	**0.934**	**0.398**
w/o SMOTE	0.912	0.356
w/o Focal Loss	0.918	0.367
w/o Contrastive	0.921	0.374
w/o MMD Detection	0.908	0.342
w/o Reweight	0.916	0.361

Removing any single component degrades performance, with the largest drops from removing MMD detection (2.8% AUC, 14.1% F1 decrease) and SMOTE augmentation (2.4% AUC, 10.6% F1 decrease). This confirms that both addressing imbalance (Stage 1) and shift (Stage 2) are essential. The relatively smaller drops from individual Stage 1 components (Focal Loss, Contrastive Learning) suggest they provide complementary benefits that partially overlap.

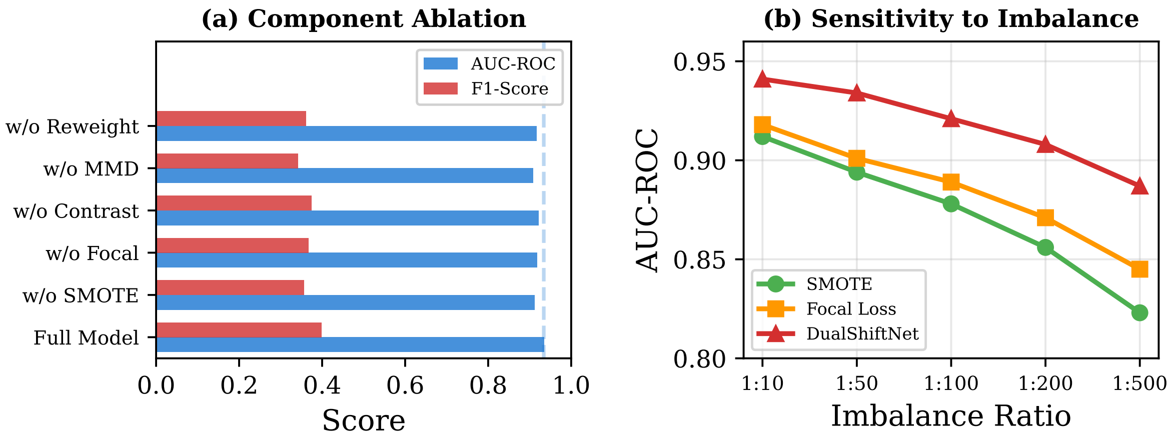

Figure 5. (a) Component ablation showing the contribution of each module to overall performance. Blue bars represent AUC-ROC scores; red bars represent F1-scores. Removing any component from Stage 1 (SMOTE, Focal, Contrastive) or Stage 2 (MMD, Reweighting) degrades performance. (b) Sensitivity analysis showing AUC-ROC under different imbalance ratios (ρ), where we vary the training minority ratio from 0.2% (1:500) to 10% (1:10). DualShiftNet maintains the highest performance across all imbalance levels.

G. Sensitivity to Imbalance Ratio

Figure 5b examines robustness to varying imbalance ratios by subsampling the Credit Card dataset. All methods degrade with increasing imbalance, but DualShiftNet maintains the smallest performance gap. At 1:500 imbalance ratio, DualShiftNet achieves 0.887 AUC compared to 0.823 for SMOTE and 0.845 for Focal Loss, demonstrating 5.0–7.8% improvement under extreme imbalance.

H. Embedding Visualization

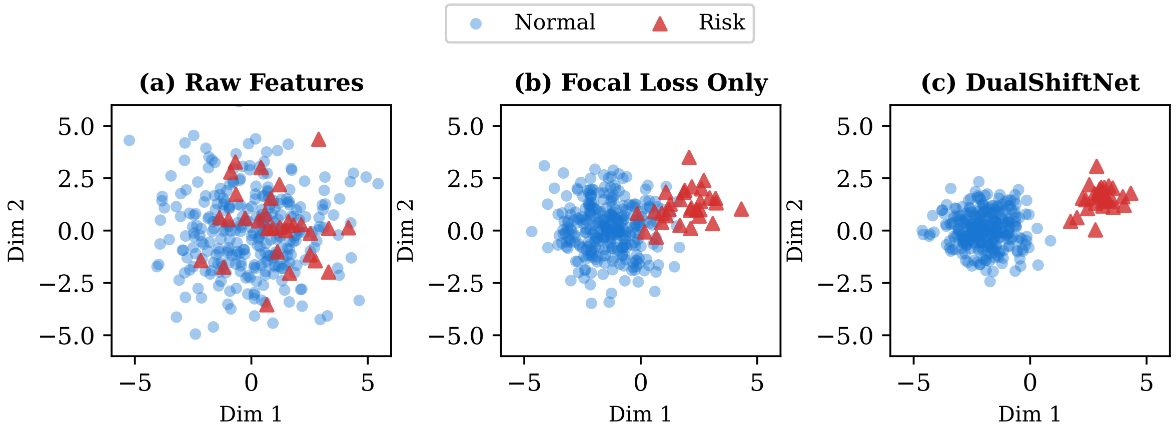

Figure 6. t-SNE visualization of embeddings on the Credit Card dataset: (a) raw features, (b) Focal Loss baseline, (c) DualShiftNet (ours). In all panels, blue points denote majority class (normal transactions); red points denote minority class (fraud). Our method produces more compact and separable class representations, with minority samples forming a tight cluster rather than being scattered among majority samples.

Figure 6 visualizes learned representations using t-SNE projections on the Credit Card dataset. In all three panels, blue points represent majority class samples (normal transactions) and red points represent minority class samples (fraud). The visualization reveals clear differences in representation quality: (a) Raw features show heavily overlapping classes with minority samples scattered sparsely among the dense majority cluster; (b) Focal Loss alone achieves moderate separation but minority samples remain dispersed without forming a coherent cluster; (c) DualShiftNet produces compact, well-separated clusters for both classes, with notably tighter grouping of minority samples. This compact minority representation is a direct result of the class-balanced contrastive learning objective, which explicitly pulls same-class samples together while pushing different-class samples apart.

VI. Discussion

Our experiments demonstrate that jointly addressing class imbalance and distribution shift provides meaningful improvements over treating them separately. The synergy arises from two factors: (1) imbalance-aware representations are more robust to shift because discriminative features are learned explicitly for minority classes rather than relying on majority-dominated statistics; and (2) shift adaptation with importance reweighting is more stable when applied to balanced contrastive embeddings that prevent weight explosion on underrepresented regions.

The adaptive threshold calibration proved particularly valuable in practice. Under high detected shift, lowering the threshold increases recall at some precision cost, providing a safety margin when model confidence is reduced. This aligns with typical business requirements where missing a risky event (false negative) is more costly than a false alarm (false positive).

Several limitations merit discussion. First, our method assumes access to unlabeled target data for MMD computation and weight estimation, which may not always be available in real-time streaming scenarios. Second, the computational overhead of contrastive learning and continuous drift monitoring may be prohibitive for extremely high-throughput applications requiring sub-millisecond latency. Third, our evaluation focused on tabular data from financial domains; extending to other modalities (text, images) or domains requires architecture modifications and additional validation.

VII. Conclusion

We presented DualShiftNet, a unified framework for business risk prediction that jointly addresses class imbalance and distribution shift. Through imbalance-aware representation learning and shift-adaptive prediction mechanisms, our approach achieves consistent improvements over methods that tackle these challenges in isolation. On three benchmark datasets, DualShiftNet improves AUC-ROC by approximately 3–4% and F1-scores by 10–22% (relative improvement) compared to the best single-challenge baselines.

The results suggest that real-world deployment of risk prediction systems should account for both challenges simultaneously rather than treating them as independent problems. Future work could explore: (1) online learning variants that continuously update representations without full retraining; (2) extension to multi-class risk categorization with hierarchical imbalance; and (3) integration with interpretability methods to explain predictions under distribution shift.

References

1. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
2. H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of Statistical Planning and Inference*, vol. 90, no. 2, pp. 227–244, 2000.
3. K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
4. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
5. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.
6. M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *Journal of Machine Learning Research*, vol. 8, no. 35, pp. 985–1005, 2007.
7. K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," *Machine Learning*, vol. 113, no. 7, pp. 4845–4901, 2024.
8. P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 18 661–18 673.
9. W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 766–785, 2021.

10. A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” *Journal of Machine Learning Research*, vol. 13, no. 25, pp. 723–773, 2012.
11. F. Kalinke, M. Heyden, G. Gntuni, E. Fouché, and K. Böhm, “Maximum mean discrepancy on exponential windows for online change detection,” *Transactions on Machine Learning Research*, 2025, iSSN 2835-8856.
12. H. He, O. Queen, T. Koker, C. Cuevas, T. Tsiligkaridis, and M. Zitnik, “Domain adaptation for time series under feature and label shifts,” in *International Conference on Machine Learning*. PMLR, 2023, pp. 12 746–12 768.
13. Ł. Korycki and B. Krawczyk, “Adversarial concept drift detection under poisoning attacks for robust data stream mining,” *Machine Learning*, vol. 112, no. 10, pp. 4013–4048, 2023.
14. S. Priya and R. A. Uthra, “Deep learning framework for handling concept drift and class imbalanced complex decision-making on streaming data,” *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 3499–3515, 2023.
15. S. Lusito, A. Pugnana, and R. Guidotti, “Solving imbalanced learning with outlier detection and features reduction,” *Machine Learning*, vol. 113, no. 8, pp. 5273–5330, 2024.
16. K. Zhang, B. Schölkopf, K. Muandet, and Z. Wang, “Domain adaptation under target and conditional shift,” in *International Conference on Machine Learning*. PMLR, 2013, pp. 819–827.
17. Machine Learning Group, Université Libre de Bruxelles, “Credit card fraud detection dataset,” Kaggle, 2018, available at: https://www.kaggle.com/mlg-ulb/creditcardfraud.
18. IEEE Computational Intelligence Society and Vesta Corporation, “IEEE-CIS fraud detection competition dataset,” Kaggle Competition, 2019, available at: https://www.kaggle.com/c/ieee-fraud-detection.
19. Kaggle, “Give me some credit dataset,” Kaggle Competition, 2011, available at: https://www.kaggle.com/c/GiveMeSomeCredit.