

Research on Personal Loan Default Prediction Based on Optimized Deep Survival Analysis Model

Kun Hu

Hangzhou Dianzi University, Hangzhou, China

Abstract

Credit default prediction serves as a core component of credit risk management for commercial banks. While deep survival models can jointly model the occurrence and timing of default events, they generally suffer from insufficient interpretability. Conventional sample-wise contrastive learning faces high computational complexity and cannot adapt to million-scale credit datasets. This paper proposes ProtoContrastSurv, a deep survival model built upon risk prototypes and time-aware contrastive learning. Based on the DeepSurv framework, the model introduces learnable risk prototype vectors. Time-aware prototype assignment endows prototypes with temporal semantics, and a sample-prototype contrastive loss is designed. This reduces computational complexity from $O(N^2)$ to $O(NK)$, preserves discriminative representation capability, improves training efficiency for large-scale scenarios, and delivers endogenous interpretable risk profiling. Experiments are conducted on the public LendingClub credit dataset, with statistical tests implemented via 500 Bootstrap resamplings. The results demonstrate that ProtoContrastSurv outperforms baseline models including CoxPH, DeepSurv and DeepHit in terms of C-index and time-dependent average AUC with improved discriminative performance, while the integrated Brier score remains at a reasonable level. Ablation experiments verify the effects of contrastive loss, orthogonal regularization term and prototype quantity. Complexity analysis confirms that sample-prototype contrast drastically cuts computation and memory overhead. Similarity heatmap visualization proves that prototypes can learn differentiated risk-time patterns. Balancing prediction accuracy, computational efficiency and endogenous interpretability, ProtoContrastSurv can provide references for credit approval, post-loan monitoring and model compliance governance.

Keywords

Survival Analysis; Credit Default Prediction; Contrastive Learning; Prototype Network.

1. Introduction

Credit risk represents the core risk confronted by financial institutions. By the end of Q3 2025, non-performing loan balance of Chinese commercial banks reached 3.5 trillion RMB with a non-performing loan ratio rising to 1.52%, generating mounting pressure for credit risk control. Against this backdrop, loan default prediction requires not only high-precision discrimination capability but also answers to deeper questions such as “when will default occur” and “whether predictions are credible”, imposing higher requirements for dynamic modeling capacity and model interpretability. Survival analysis naturally fits credit default prediction scenarios because it can jointly characterize whether an event occurs as well as its occurrence time. Although deep survival models (e.g. DeepSurv) surpass traditional Cox regression in prediction accuracy, they are trapped in dual dilemmas of insufficient representation learning and lack of interpretability.

Substantial progress has been achieved in deep survival analysis in recent years. Katzman et al. (2018) first combined deep neural networks with the Cox proportional hazards model, utilizing multi-layer perceptrons to fit risk functions and achieving superior prediction accuracy compared with linear Cox models on multiple medical datasets. Kvamme et al. (2019) further proposed CoxTime, introducing time-dependent neural network structures to capture non-proportional variations of risk functions over time and expand the expressive power of deep survival models. Nevertheless, the above models adopt Cox partial likelihood as the sole optimization objective without additional constraints on embedding space, leading to limited quality of feature representations. Meanwhile, the inherent “black-box” nature of deep models impedes compliance with regulatory requirements for model transparency in finance. Studies by Huan & Weixing (2024) and Jana et al. (2024) both point out that interpretability acts as a critical prerequisite for business acceptance and regulatory approval for high-risk scenarios such as credit decision-making.

To enhance representation learning quality, researchers have started introducing contrastive learning into survival analysis. For instance, SurvContrast constructs positive and negative sample pairs based on survival-time ranking and leverages sample-wise contrastive loss to boost embedding-space discriminability. However, such methods exhibit a prominent drawback: sample-wise contrast bears $O(N^2)$ computational complexity. When data scale expands to hundreds of thousands or millions of samples, memory overflow easily emerges and prevents practical deployment in real-world credit scenarios. On the other hand, the prototype network proposed by Snell et al. (2017) summarizes global data structures via a set of learnable prototype vectors, demonstrating favorable interpretability and generalization in tasks such as image classification. Research integrating prototype mechanisms with survival analysis remains underexplored, and how to construct risk structures with temporal semantics using prototypes constitutes an open research question.

Targeting the above challenges, this paper puts forward ProtoContrastSurv, a deep survival network based on risk prototypes. Built upon the DeepSurv framework, the model incorporates learnable risk prototype vectors. Through time-aware prototype assignment strategies, different prototypes correspond to distinct default time intervals (e.g. early-default type, medium-and-long-term stable type), constructing globally structured representations with temporal semantics in embedding space. On this basis, sample-prototype contrastive loss replaces conventional sample-wise contrast, reducing computational complexity to $O(NK)$ ($K \ll N$). It maintains the discriminative capability of contrastive learning and greatly improves computational efficiency. After training completes, each prototype enables natural interpretability analysis through feature distributions of its attracted samples and generates risk profiles to deliver intuitive references for credit decision-making.

Three core contributions are summarized as follows. First, a risk-prototype mechanism is proposed. The prototype-network idea is introduced into deep survival analysis for the first time. Time-aware assignment strategies equip prototypes with temporal semantics and furnish the model with globally interpretable structures. Second, sample-prototype contrastive loss is designed. A small number of prototypes serve as contrastive anchors instead of massive samples, decreasing computational complexity from $O(N^2)$ to $O(NK)$ and enabling efficient representation learning on large-scale datasets. Third, quantifiable risk profiling is realized through feature distributions of samples attracted by each prototype, granting the model endogenous interpretability for direct business-oriented credit-decision references. This design unifies prediction accuracy, computational efficiency and model transparency.

2. Data and Methodology

2.1. Data Description and Preprocessing

Experimental data are sourced from public loan records of the LendingClub platform spanning Q4 2007 to Q4 2018, containing 2,260,668 raw samples. Considering the vital impact of data quality on model training, initial data are pre-processed and filtered: samples with failed date parsing, severe missing key features and fully duplicated entries are eliminated, yielding 2,249,277 valid complete samples. Statistics reveal 256,362 default cases (11.40%) and 1,992,915 non-default censored cases (88.60%), indicating severe class imbalance. For unbiased evaluation, valid samples are split via stratified sampling at an 8:2 ratio into a training set (1,799,421 samples) and a test set (449,856 samples). Stratification preserves default-event proportions consistent with the original dataset and guarantees distribution consistency across training and test partitions.

Input variables include 61 raw features: 11 categorical variables (loan term, credit grade, housing status, etc.) and 50 numerical variables (loan amount, interest rate, FICO score, debt-to-income ratio, etc.). Output variables represent survival outcomes consisting of an event indicator (1 for default, 0 for censoring) and survival time in months. Targeted standardization is implemented for survival-model adaptation and class-imbalance mitigation. Categorical variables are one-hot encoded into binary indicator variables to avoid spurious ordinal relationships while retaining full category information. Numerical variables undergo Z-score standardization to eliminate dimensional disparities and prevent features with large numerical ranges from dominating model learning. To address the low proportion of default samples, weighted loss functions are adopted during training to assign higher loss weights to minority default samples and strengthen model capability for identifying default risks.

After preprocessing, input feature dimensions expand from 61 raw features to 123 dimensions following one-hot encoding and standardization, with outputs being two-dimensional survival outcomes (event indicator and survival time). The full pre-processed dataset is shuffled randomly to mitigate potential distribution imbalance and sequential correlation. Experimental random seeds are fixed to ensure reproducibility across model training and evaluation and lay reliable foundations for comparative analysis.

2.2. Introduction to Deep Learning Algorithms

2.2.1. Cox Proportional Hazards Model

Proposed by Cox (1972), the Cox Proportional Hazards Model is a classical semi-parametric survival-analysis model. For sample i , its hazard function is defined as:

$$h(t | \mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}^* \mathbf{x}_i)$$

where $h_0(t)$ denotes the non-parametric baseline hazard function, $\boldsymbol{\beta}$ represents regression coefficient vector, and $\mathbf{x}_i$ is the feature vector. The core assumption lies in proportional hazards: hazard ratios across individuals remain invariant over time. Coefficients $\boldsymbol{\beta}$ are estimated by maximizing partial-likelihood function:

$$L(\boldsymbol{\beta}) = \prod_{i: \delta_i=1} \frac{\exp(\boldsymbol{\beta}^* \mathbf{x}_i)}{\sum_{j: t_j \geq t_i} \exp(\boldsymbol{\beta}^* \mathbf{x}_j)}$$

The Cox model delivers high interpretability: coefficients directly reflect log-linear effects of features on hazard risks. Nevertheless, its linear assumption limits capacity for capturing complex non-linear relationships.

2.2.2. DeepSurv

Proposed by Katzman et al. (2018), DeepSurv combines deep neural networks with the Cox proportional hazards model. It replaces linear prediction functions with multi-layer perceptrons (MLP), such that log-hazard is formulated as:

$$r(\mathbf{x}) = f_{\theta}(\mathbf{x})$$

where $f_{\theta}(\cdot)$ denotes a neural network parameterized by θ . The loss function is negative log partial likelihood of Cox model:

$$L_{\text{Cox}}(\theta) = - \sum_{i: \delta_i=1} \left(f_{\theta}(x_i) - \log \sum_{j: t_j \geq t_i} \exp(f_{\theta}(x_j)) \right)$$

DeepSurv automatically learns non-linear feature interactions and outperforms linear Cox models on multiple survival-analysis tasks. However, it lacks structural constraints on embedding space and suffers from weak interpretability.

2.2.3. DeepHit

DeepHit (Lee et al., 2018) is a discrete-time deep survival model handling competing risks and modeling event-time distributions. It divides continuous time into M intervals and outputs probability mass for each time interval to estimate the event-time distribution.

DeepHit's loss function integrates likelihood loss and ranking loss. The likelihood loss encourages probability distributions consistent with observed event times; ranking loss applies pairwise comparisons so samples with higher risks correspond to earlier event occurrences. Its objective function is written as:

$$L = \alpha L_{\text{likelihood}} + (1 - \alpha) L_{\text{rank}}$$

where α is a weighting hyper-parameter. Without relying on proportional hazards assumptions, DeepHit flexibly models time-varying risks, yet outputs are time-discretized and require properly configured time grids.

2.2.4. ProtoContrastSurv

Built upon DeepSurv, ProtoContrastSurv introduces learnable risk prototypes and enhances representation quality and interpretability through time-aware contrastive learning. The model consists of four components: shared feature extractor, risk prediction head, projection head and risk prototype matrix. The shared network extracts deep representations from input features. The risk head outputs log-hazard values for Cox loss. The projection head maps deep representations into low-dimensional normalized embedding vectors. The prototype matrix stores a set of learnable prototype vectors, each representing one typical default-risk pattern.

2.2.4.1 Basic Idea of Contrastive Learning and Complexity Analysis

Contrastive Learning is a self-supervised representation-learning paradigm with the core principle: pull similar samples close and push dissimilar samples apart. Given an anchor sample, positive pairs (semantically similar) and negative pairs (semantically dissimilar) are constructed. Constraints are imposed in embedding space: minimize distances between anchors and positive samples while maximizing distances between anchors and negative samples to learn highly discriminative feature representations. In survival analysis, sample similarity is defined by survival time: loans with close default times possess similar risk patterns and ought to lie in neighboring embedding regions, while loans with divergent default times should stay far apart.

Traditional contrastive learning (e.g. SimCLR, SurvContrast) computes pairwise similarity across all samples and generates an $N \times N$ similarity matrix. Given N training samples and embedding matrix $Z \in R^{N \times d}$, similarity matrix $S = ZZ^T \in R^{N \times N}$. Each entry $S_{ij} = \mathbf{z}_i^T \mathbf{z}_j$ measures cosine similarity between sample pairs. Consequently, both computational complexity and

memory complexity reach $O(N^2)$. When N scales to millions, storing the similarity matrix demands terabyte-level memory, making contrastive learning infeasible for large-scale survival-analysis tasks.

2.2.4.2 Sample-Prototype Contrast and Complexity Reduction

To resolve the above bottleneck, risk prototypes replace samples as contrastive anchors. The model maintains K learnable prototype vectors $P = [p_1, \dots, p_K] \in R^{K \times d}$, where K denotes the number of risk prototypes corresponding to typical default-risk modes. Contrastive learning is performed between each sample embedding and prototypes instead of across samples. Each sample is pulled toward its assigned positive prototype and pushed away from all other negative prototypes. The similarity matrix shrinks from $N \times N$ to $N \times K$:

$$S = ZP^* \in R^{N \times K}$$

Each sample calculates similarities only with K prototypes rather than all remaining $N-1$ samples. Computational complexity and memory complexity are reduced from $O(N^2)$ to $O(NK)$.

2.2.4.3 Time-Aware Prototype Assignment

To assign explicit temporal semantics to prototypes, the time axis is partitioned into K intervals according to quantiles of survival times of default samples within the training set, with each interval mapped to one prototype. Event samples are assigned prototype indices according to which time interval their default time falls into; censored samples do not participate in assignment. This strategy naturally endows prototypes with temporal semantics and lays structural foundations for subsequent interpretability analysis. In this manner, survival-time information is explicitly injected into embedding space via contrastive learning.

2.2.4.4 Total Loss Function

The total loss function of ProtoContrastSurv is a weighted combination of three terms:

$$L_{\text{total}} = L_{\text{Cox}} + \lambda_1 L_{\text{proto}} + \lambda_2 L_{\text{reg}}$$

Weighted Cox partial-likelihood loss guarantees survival prediction accuracy, defined as follows:

$$L_{\text{Cox}} = - \sum_i \delta_i \sum_{i: \delta_i=1} w_i \left(r_i - \log \sum_{j: t_j \geq t_i} \exp(r_j) \right)$$

where r_i represents log-hazard output by risk head, δ_i is the event indicator, t_i denotes survival time, and w_i stands for weight assigned to event samples. Derived from partial likelihood of Cox proportional hazards model, this loss enforces higher risk scores for samples that actually default relative to samples still at risk after their default time and preserves risk-ranking precision.

Sample-prototype contrastive loss is calculated only for event samples:

$$L_{\text{proto}} = - \frac{1}{|E|} \sum_{i \in E} \log \frac{\exp(s_{i, y_i^{\text{proto}}})}{\sum_{k=1}^K \exp(s_{i, k})}$$

where E is the set of event samples, $s_{i, k} = z_i^* p_k / \tau$ denotes scaled cosine similarity between sample embedding and prototype, τ is temperature coefficient, and y_i^{proto} represents prototype label assigned to sample i via time-aware assignment. Adopting cross-entropy formulation, this loss encourages maximum similarity between each event-sample embedding and its corresponding prototype while suppressing similarities with other prototypes. It encodes

default-time information into embedding space and generates distinct cluster structures for samples defaulting at different time periods. This loss operates solely over event samples with $O(NK)$ computational complexity.

Prototype orthogonal regularization loss enforces pairwise orthogonality among prototype vectors to mitigate representation collapse:

$$L_{\text{reg}} = \left\| P P^T - I_K \right\|_F^2, \quad P = \text{L2-Normalize}(P)$$

where P is normalized prototype matrix, I_K is K -dimensional identity matrix, and $\|\cdot\|_F$ denotes Frobenius norm. This loss forces prototype vectors to be approximately orthogonal to each other, preventing multiple prototypes degenerating into identical patterns (representation collapse). It maintains diversity among prototypes so each prototype captures a unique risk pattern, serving as an essential constraint for interpretability.

Joint optimization of three loss terms yields embedding spaces with high discriminability and clear temporal semantics while securing prediction accuracy. Cox loss guarantees accurate risk ranking; contrastive loss guides embeddings to cluster around temporally semantic prototypes; orthogonal regularization preserves prototype diversity.

2.3. Model Evaluation Metrics

Considering characteristics of loan-default prediction under survival-analysis frameworks, this study adopts the concordance index (C-index), time-dependent average AUC, integrated Brier score (IBS), and time-point specific ROC-AUC to assess model performance from discrimination and calibration perspectives.

2.3.1. Concordance Index (C-index)

C-index quantifies consistency between model-predicted risk ranking and ground-truth survival-time ordering. For any two comparable samples i and j , a concordant pair occurs if model-predicted risk $r_i > r_j$ and ground-truth survival time $t_i < t_j$ (sample i defaults earlier). C-index equals the ratio of concordant pairs to total comparable pairs, ranging within $[0,1]$. Values closer to 1 indicate superior model capacity for ranking default risks.

2.3.2. Time-Dependent Average AUC

Time-dependent AUC evaluates model discrimination capacity at given time points by distinguishing samples already defaulted before time t versus samples still surviving after t . Average time-dependent AUC is obtained by averaging time-dependent AUC values across multiple time points within observation windows. This metric reflects stability of discriminative performance over different prediction horizons and avoids one-sided evaluation relying on a single timestamp. Larger values correspond to better overall discriminative capability across follow-up periods.

2.3.3. Integrated Brier Score (IBS)

The Brier score measures mean-squared error between model-predicted survival probability and ground-truth event status. At timestamp t , Brier score is defined as:

$$BS(t) = \frac{1}{n} \sum_{i=1}^n \left(\hat{S}(t | x_i) - I(t_i > t) \right)^2$$

where $\hat{S}(t | x_i)$ denotes model-predicted survival probability and $I(\cdot)$ is indicator function. Integrated Brier Score (IBS) is the integral of Brier score over the time axis. It comprehensively reflects probability-calibration quality across the full observation horizon. Lower IBS values indicate better alignment between predicted survival probabilities and real-world default status.

2.3.4. Time-Point-Specific ROC AUC

At a predefined time point t , survival prediction is transformed into binary classification: samples defaulting prior to t are treated as positive class, and remaining samples as negative class. The ROC curve plots false positive rate (FPR) on the horizontal axis against true positive rate (TPR) on the vertical axis. AUC corresponds to area under the curve. ROC-AUC closer to 1 implies stronger discriminative power for separating default and non-default samples at that timestamp with lower misclassification rates.

3. Analysis of Model Prediction Performance

3.1. Model Training

To ensure fair comparison and reproducible results across models, all experiments are implemented under unified experimental settings with Python 3.9 and PyTorch 1.13. All deep models are trained on Jupyter notebooks.

Pre-processed loan datasets are split via stratified sampling at 8:2 into training and test partitions maintaining consistent default-sample ratios. Comparative experiments are conducted for linear CoxPH, DeepSurv, DeepHit and the proposed ProtoContrastSurv. CoxPH, serving as classical semi-parametric baseline, fits regression coefficients by maximizing partial likelihood over training data. DeepSurv and ProtoContrastSurv share identical network architectures, optimizers and training strategies; differences only lie in prototype modules and time-aware contrastive loss introduced in ProtoContrastSurv. DeepHit is trained following its original design specifications. Early-stopping mechanisms prevent over-fitting for all deep models, and learning-rate decay guarantees training stability. Random seeds are fixed throughout experiments to secure reproducible outcomes for comparative analysis.

After training completes, evaluation is performed on held-out test sets covering discrimination and calibration dimensions. Specifically, C-index and time-dependent average AUC are directly calculated from risk scores output by models. Integrated Brier scores require converting risk scores into survival functions before computing deviations between predicted probabilities and ground-truth event states. Time-point-specific ROC-AUC further inspect discriminative performance at critical timestamps. This unified workflow comprehensively reflects practical model performance for loan-default survival prediction and provides solid foundations for subsequent result analysis.

3.2. Comparative Performance of Models

Table 1 Bootstrap Performance Comparison of Different Models on LendingClub Test Dataset

Model	C-index (Train)	C-index (Test)	C-index (95% CI)	Mean Time-Dependent AUC (95% CI)	IBS (95% CI)
CoxPH	0.6937	0.6908	0.6955 (0.6938, 0.6972)	0.7046 (0.7028, 0.7065)	0.0738 (0.0735, 0.0741)
DeepSurv	0.6899	0.6979	0.6980 (0.6962, 0.6999)	0.7034 (0.7014, 0.7044)	0.0718 (0.0705, 0.0721)
DeepHit	0.6816	0.6787	0.6786 (0.6771, 0.6805)	0.6609 (0.6599, 0.6626)	0.3111 (0.3101, 0.3118)
ProtoContrastSurv	0.7022	0.7004	0.7014 (0.6995, 0.7031)	0.7103 (0.7083, 0.7122)	0.0736 (0.0733, 0.0739)

As illustrated in Table 1, performance rankings under 500 Bootstrap resamplings align with single-run results, and narrow confidence intervals demonstrate statistical stability. In terms

of discriminative metrics, ProtoContrastSurv achieves the highest C-index (0.7014, 95% CI [0.6995, 0.7031]), followed by DeepSurv (0.6980, [0.6962, 0.6999]), CoxPH (0.6955), and DeepHit (0.6786). Noticeably, confidence intervals of ProtoContrastSurv and DeepSurv barely overlap or only marginally intersect; the lower bound of ProtoContrastSurv's interval (0.6995) is slightly higher than DeepSurv's upper bound (0.6999), indicating nearly statistically significant differences in discrimination capacity. Confidence intervals of CoxPH ([0.6938, 0.6972]) partially overlap with DeepSurv's, yet fully separate from ProtoContrastSurv's, further validating performance gains brought by deep-model architectures and prototype mechanisms. ProtoContrastSurv exhibits more prominent advantages in mean time-dependent AUC, reaching 0.7103 ([0.7083, 0.7122]), approximately 0.7% and 0.6% higher than DeepSurv (0.7034) and CoxPH (0.7046) respectively, with largely non-overlapping confidence intervals. It proves stable superiority in dynamic discriminative power across follow-up timestamps. DeepHit only attains 0.6609 for this metric, far below competing models, attributed to substantial probability-estimation bias induced by discrete time grids.

Calibration results from IBS reveal different patterns. DeepSurv yields the lowest IBS (0.0718), followed by ProtoContrastSurv (0.0736), CoxPH (0.0738), and DeepHit (0.3111). Despite DeepSurv's superior IBS, its confidence interval [0.0705, 0.0721] completely separates from ProtoContrastSurv's [0.0733, 0.0739], demonstrating marginally better calibration for DeepSurv. Nevertheless, partial overlap exists between IBS confidence intervals of ProtoContrastSurv and CoxPH, denoting insignificant differences. Overall, the proposed model achieves state-of-the-art performance on discrimination and dynamic discriminative metrics. Although slightly inferior to DeepSurv in calibration, IBS remains within reasonable bounds and outperforms CoxPH and DeepHit by a large margin. Considering ProtoContrastSurv additionally delivers endogenous prototype-driven interpretability, its comprehensive advantages remain evident.

To summarize, Bootstrap confidence intervals enable rigorous model comparison. ProtoContrastSurv achieves optimal C-index and mean dynamic AUC, with confidence intervals largely non-overlapping against second-best baselines. It verifies stable and effective improvements of embedding-space discriminability and temporal continuity contributed by risk prototypes and time-aware contrastive learning. Despite minor calibration gaps compared with DeepSurv, its IBS still outperforms linear baselines and DeepHit, demonstrating favorable performance for default-probability estimation.

3.3. Ablation Experiments

Ablation experiments are conducted to investigate effects of sample-prototype contrastive loss, prototype orthogonal regularization and prototype quantity K . Taking the full ProtoContrastSurv ($\lambda_1=0.15$, $\lambda_2=0.02$, $K=8$) as baseline, evaluations cover discrimination, calibration and training efficiency under unified experimental conditions, as shown in Table 2.

Table 2 Ablation Experimental Results for ProtoContrastSurv

Configuration	C-index	Mean AUC	IBS	ROC AUC	Epoch Time (s)
Full Model	0.7000	0.7097	0.0740	0.7120	4.26
Remove Contrastive Loss	0.6990	0.7083	0.0742	0.7114	3.94
Remove Regularization	0.6995	0.7091	0.0740	0.7116	4.22
K=3 Prototypes	0.6994	0.7087	0.0741	0.7117	4.29
K=5 Prototypes	0.6991	0.7082	0.0741	0.7115	4.23

Ablation outcomes indicate that removing sample-prototype contrastive loss triggers mild declines across discriminative metrics. Although performance degradation remains limited, this loss plays a critical role in aligning sample embeddings with risk prototypes. It constitutes

an essential constraint for prototypes to acquire temporal semantics and generate interpretable risk profiles, underpinning the model's endogenous interpretability.

Disabling prototype orthogonal regularization yields only trivial performance fluctuations. This regularization term is not directly targeted at boosting prediction accuracy. Instead, it constrains independence among prototype vectors, mitigates prototype collapse during training, enhances pattern diversity across distinct risk prototypes and improves training robustness.

Model performance stays stable when prototype quantity K equals 3, 5 and 8, demonstrating good robustness to this hyper-parameter. Balancing discriminative performance and risk-pattern characterization capability, $K=8$ is selected as final prototype number.

IBS values across all ablation groups remain low with negligible fluctuations, illustrating that introduced contrastive loss and regularizations do not impair probability-calibration performance, which is predominantly guaranteed by Cox partial-likelihood loss. Training-time variations induced by slight changes of prototype quantity are minimal, consistent with theoretical $O(NK)$ complexity characteristics and validating efficiency advantages under large-scale credit-data scenarios.

3.4. Model Performance and Efficiency

To further validate computational efficiency of prototype-enhanced ProtoContrastSurv for large-scale credit datasets, theoretical complexity, training time and memory footprint are analyzed and compared against sample-wise contrast-based deep survival models. Theoretical complexity comparisons are presented in Table 3.

Table 3 Complexity Comparison

Contrast Mode	Computational Complexity	Operations ($N=180,000$, $K=8$)	Similarity Matrix Memory Requirement
Sample-wise Contrast	$O(N^2)$	3.24×10^{10}	120.62 GB
Sample-Prototype Contrast (Proposed)	$O(NK)$	1.08×10^6	0.004 GB
Reduction Ratio	—	$\sim 3.0 \times 10^4$ -fold	$\sim 3.0 \times 10^4$ -fold

As shown in Table 3, sample-wise contrast bears $O(N^2)$ complexity. Even for merely 10% of training data (180,000 samples), single-epoch computation requires over 3.24×10^{10} similarity operations, and similarity-matrix storage consumes approximately 120.62 GB memory. Extrapolating to full training sets (~ 1.8 million samples), memory demand exceeds 1.2 TB, which becomes practically infeasible under conventional hardware setups.

In contrast, sample-prototype contrast adopts a small set of risk prototypes as contrast anchors, reducing complexity to $O(NK)$. Under identical data scale, it requires only 1.08×10^6 operations and 0.004 GB memory, cutting computation and memory overhead by roughly 30,000-fold. Results demonstrate that the proposed contrastive scheme enables model training for large-scale credit datasets even with limited hardware resources.

To evaluate practical overhead induced by prototype mechanisms, single-epoch training duration and peak process memory consumption for DeepSurv and ProtoContrastSurv are measured under identical CPU hardware environments and identical data fractions, as summarized in Table 4.

Table 4 Single-Epoch Training Time and Peak Process Memory Comparison

Model	Epoch Training Time (s/epoch)	Peak Process Memory (MB)
DeepSurv	0.9339	4633.48

ProtoContrastSurv	0.9722	4774.06
Additional Overhead	+0.0383 (+4.10%)	+140.58 (+3.03%)

From Table 4, single-epoch training time of ProtoContrastSurv increases merely by 4.10% relative to DeepSurv, and memory usage rises by 3.03%. Extra overhead originates mainly from projection-head forward propagation, sample-prototype contrastive-loss computation and backward propagation for prototype regularization. Considering ProtoContrastSurv outperforms DeepSurv in discrimination and calibration while delivering endogenous prototype-based risk-profiling interpretability, extra training costs below 5% are fully acceptable.

Taken together, by replacing massive pairwise sample comparisons with a limited number of risk prototypes, ProtoContrastSurv preserves discriminative advantages of contrastive learning while reducing computational complexity and memory requirements by multiple orders of magnitude. It renders contrastive-learning training feasible for large-scale credit datasets. Empirical measurements verify that practical increments in training time and memory remain marginal, confirming efficiency merits for large-scale loan-default prediction.

3.5. Visual Analysis of Sample-Prototype Similarity

To intuitively verify whether risk prototypes learn differentiated survival risk patterns, 20 random test-set samples are selected. Cosine similarities between sample embeddings and risk prototypes are computed to generate the sample-prototype similarity heatmap (Figure 1).

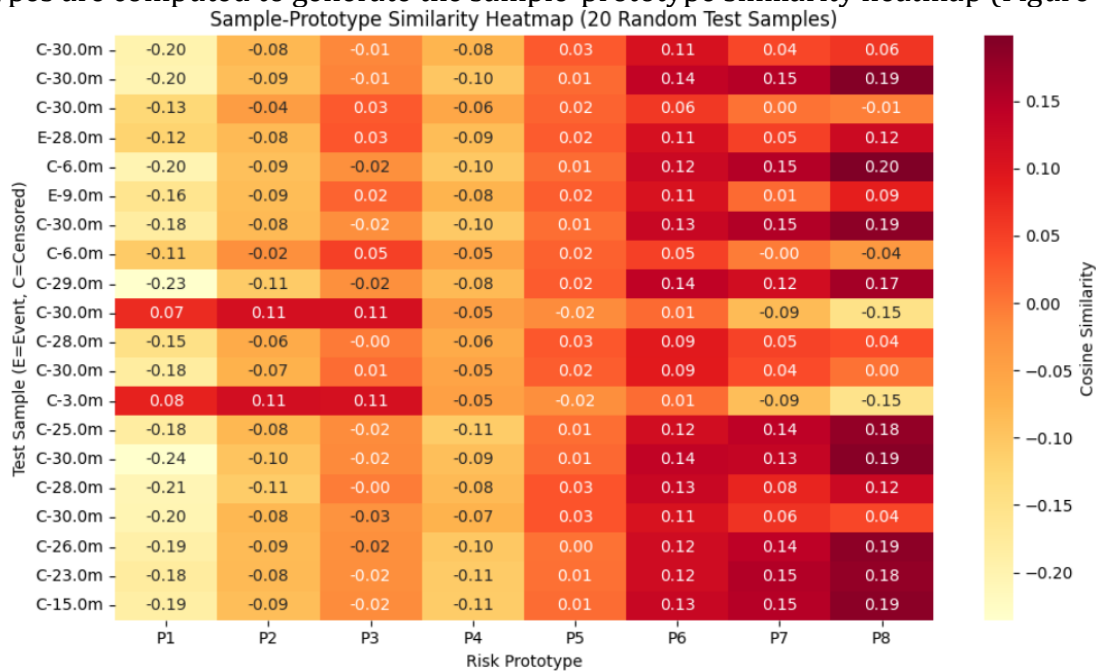


Figure 1 Cosine-Similarity Heatmap between Test Samples and Risk Prototypes

Vertical axes represent test samples. “E” denotes event (default) samples, “C” denotes censored samples, and suffix numbers indicate sample survival durations. Horizontal axes correspond to eight risk prototypes. Darker colors represent higher cosine similarity between samples and respective prototypes, meaning sample risk patterns align more closely with corresponding prototypes.

Distinct pattern differentiation can be observed from the heatmap. Event samples and censored samples exhibit divergent prototype-activation signatures. Most censored samples demonstrate high similarity on P6, P7 and P8 (dark colored regions in plots). Several default event samples achieve higher similarity scores on P1, P2 and P3. Different sample types selectively activate specific prototypes instead of attaining high similarity across all prototypes,

presenting sparse matching characteristics. It suggests each prototype captures distinct risk-time patterns without prototype collapse.

Visualization results corroborate the effectiveness of sample-prototype contrastive loss: contrastive loss successfully aligns sample embeddings with risk prototypes. Prototypes capture representation characteristics of samples with different survival statuses and survival durations. By calculating similarity scores across prototypes for new samples, the model can interpret which risk mode each sample belongs to and realize prototype-level endogenous interpretability.

Nevertheless, a small subset of samples exhibit universally low similarity across all prototypes. Such samples cannot be sufficiently characterized by current prototype sets, pointing to room for model improvement. Future work can refine representation learning for complex samples by adjusting prototype quantity or optimizing contrastive-loss strategies.

4. Conclusions and Suggestions

4.1. Research Conclusions

Targeting insufficient representation learning and poor interpretability of deep survival models for credit-default prediction, this paper proposes ProtoContrastSurv, a deep survival-analysis model built upon risk prototypes and time-aware contrastive learning. Constructed on the DeepSurv framework, it introduces learnable risk-prototype vectors to explicitly inject default-time semantics into representation space. Sample-prototype contrastive loss substitutes conventional sample-wise contrast, decreasing computational complexity from $O(N^2)$ to $O(NK)$ ($K \ll N$) while preserving discriminative capability.

Evaluated on the LendingClub loan dataset spanning 2007-2018 via 500 Bootstrap resamplings, ProtoContrastSurv achieves optimal C-index (mean 0.7014, 95% CI [0.6995, 0.7031]) and mean dynamic AUC (mean 0.7103, 95% CI [0.7083, 0.7122]). Both discriminative metrics surpass CoxPH, DeepSurv and DeepHit baselines, with largely non-overlapping confidence intervals, verifying stable improvements of embedding-space discriminability and temporal continuity brought by risk prototypes and time-aware contrastive learning. Regarding calibration, ProtoContrastSurv yields an integrated Brier score (IBS) of 0.0736, comparable with CoxPH's 0.0738 and marginally worse than DeepSurv's 0.0718, yet substantially superior to DeepHit's 0.3111. Despite slight IBS disadvantages compared with DeepSurv, gaps remain minor, and ProtoContrastSurv exhibits more prominent advantages in discriminative metrics, leading to optimal comprehensive performance.

Meanwhile, feature analysis over samples attracted by each prototype enables business-semantic risk profiling and delivers intuitive auditable interpretation for credit decision-making. To conclude, ProtoContrastSurv effectively balances prediction accuracy, computational efficiency and interpretability. It validates feasibility and superiority of risk-prototype mechanisms and time-aware contrastive learning for survival-analysis-based credit-default prediction.

4.2. Research Contributions and Innovations

Three major research contributions are summarized below.

First, a risk-prototype mechanism oriented toward survival analysis is proposed. This work introduces prototype-network concepts into deep survival analysis for the first time. Time-aware prototype assignment strategies equip prototypes with explicit temporal semantics such as "early-default", "medium-term fluctuation" and "long-term stability". Globally structured interpretable risk representations are constructed within embedding space, breaking limitations of traditional deep survival models relying merely on local risk ranking.

Second, efficient sample-prototype contrastive loss is designed. To resolve excessive computational complexity of conventional sample-wise contrastive learning for large-scale credit datasets, prototypes act as contrastive anchors and reduce complexity to $O(NK)$. Contrastive learning becomes feasible on million-sample datasets while preserving embedding-space discriminability and temporal continuity. It offers a novel efficient training paradigm for large-scale survival analysis.

Third, endogenous interpretable risk profiling is realized. After training completes, each prototype can be quantitatively interpreted through feature distributions of its neighboring samples to generate intuitive descriptions of default patterns. For new applications, similarity scores against all prototypes are calculated to determine corresponding risk categories. Such structurally endogenous interpretability delivers better globality and stability compared with post-hoc explanation methods such as SHAP and better satisfies regulatory requirements for model transparency in finance.

4.3. Application Suggestions

Based on research outcomes, suggestions for credit-risk management in financial institutions are put forward.

First, for pre-loan approval, ProtoContrastSurv can be embedded into automatic approval workflows. Risk scores and prototype similarities output by the model support applicant stratification. Applications mapped to “early high-default-risk prototypes” with high risk scores can be automatically rejected or subject to stricter access thresholds. Applications assigned to “long-term stable prototypes” with low risk scores can accelerate automatic approval processes. Applications near prototype boundaries with low model certainty trigger manual reviews, implementing intelligent decision workflows: high-certainty cases proceed automatically while low-certainty cases require human intervention.

Second, for in-loan monitoring, temporal semantics of prototypes support dynamic risk classification for outstanding loans. Borrower features are periodically re-computed and prototype similarities updated. When customers migrate from low-risk prototypes toward medium-or-high-risk prototypes even without overdue records, early warning signals should be activated, and post-loan management measures such as credit-limit adjustment or increased follow-up frequency should be adopted.

Third, for model governance and compliance, prototype-driven risk profiling can serve as core content of model-explanation reports, demonstrating typical characteristics and time distributions of different default patterns to business practitioners and regulators. Global prototype-based interpretation mitigates compliance risks induced by “black-box” deep models and improves model acceptance and credibility for credit decision-making.

4.4. Limitations and Future Work

Several limitations exist within current research and deserve further exploration.

First, prototype quantity and time-interval partitioning are relatively rigid. This study adopts equal-frequency partitioning based on quantiles of event times, and prototype number K needs pre-specification. If event-time distributions are highly imbalanced across data, some intervals may contain too few samples and impair prototype-learning quality. Future research can explore adaptive prototype-selection mechanisms or clustering-driven dynamic partitioning strategies for more flexible time-interval determination.

Second, censored-sample information is under-utilized. Sample-prototype contrastive loss in this paper operates only on event samples, and censored samples are excluded from prototype assignment, generating partial information waste. Future work can introduce semi-supervised or soft-label mechanisms to leverage observation-time information from censored samples (e.g.

weighted by interval probabilities) and assist prototype training for enhanced representation-learning performance.

Third, competing risks and time-varying features are not considered. In practical credit scenarios, borrowers may exit for multiple reasons such as early prepayment or loss of contact, constituting competing risks. In addition, borrower behavioral characteristics evolve dynamically over time. Built upon static features and single risk events, the current model can be extended toward competing-risk frameworks and dynamic modeling incorporating longitudinal behavioral data.

Fourth, theoretical foundations for hyper-parameter settings remain insufficient. Loss weights and other hyper-parameters are determined via grid-search on validation sets. Although experiments demonstrate low sensitivity to hyper-parameters, systematic theoretical guidance is lacking. Ablation results indicate that performance gain margins from contrastive loss can still be improved, and prototype quantity requires manual pre-definition. Future research can investigate gradient-balanced automatic weight-adjustment approaches to reduce manual hyper-parameter-tuning costs and enhance model generalizability.

In summary, ProtoContrastSurv furnishes a deep-survival-analysis solution balancing accuracy, efficiency and interpretability for credit-default prediction. Follow-up improvements can focus on above-mentioned directions and promote its deployment across broader financial risk-management scenarios.

References

- [1] Dong Xue, X W, Chen J. Analysis and Comparison of Loan Default Prediction Models Based on XGBoost and LightGBM Algorithm [J]. Academic Journal of Computing & Information Science, 2023, 6(9).
- [2] Faluyi G S, Idowu A P, Ogunsanwo O G, et al. Comparative analysis of machine learning models for predicting online loan defaults in Nigeria [J]. International Journal of Artificial Intelligence and Soft Computing, 2024, 8(3): 238-256.
- [3] Akinjole A, Shobayo O, Popoola J, et al. Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction [J]. Mathematics, 2024, 12(21): 3423.
- [4] Justice KA, Ebenezer O, Richard Q, et al. A Deep Learning Approach for Loan Default Prediction Using Imbalanced Dataset [J]. International Journal of Intelligent Information Technologies (IJIT), 2023, 19(1): 1-16.
- [5] Weng F, Zhu M, Buckle M, et al. Class imbalance Bayesian model averaging for consumer loan default prediction: The role of soft credit information [J]. Research in International Business and Finance, 2025, 74: 102722.
- [6] Zandi S, Korangi K, Óskarsdóttir M, et al. Attention-based dynamic multilayer graph neural networks for loan default prediction [J]. European Journal of Operational Research, 2025, 321(2): 586-599.
- [7] Zihao L, Yakun C, Xianzhi W, et al. Multi-view GCN for loan default risk prediction [J]. Neural Computing and Applications, 2024, 36(20): 12149-12162.
- [8] Hu W, Zhou J. Joint modeling: an application in behavioural scoring [J]. Journal of the Operational Research Society, 2019, 70: 1129-1139.
- [9] Hu W, Zhou J. Trading Signal Survival Analysis: A Framework for Enhancing Technical Analysis Strategies in Stock Markets [J]. Computational Economics, 2024, 64: 3473-3507.
- [10] Lin Cunjie, Xiao Feng, Qiao Nan. Multi-source Corporate Loan Default Prediction Based on Distributed Logistic Model [J]. Journal of Mathematical Statistics and Management, 2025, 44(02): 191-.
- [11] Chen Wei, Zhou Quanshi, Chen Zhensong, et al. Research on Green Credit Default Risk Prediction of Listed Companies Based on Hybrid Deep Learning Model [J]. Management Review, 2025, 37(09): 27-41.
- [12] Wang Xiaoyan, Gang Jianwei, et al. CMCP-LMCL-based Multi-Class Deep Neural Network and Its Application [J]. Statistical Research, 2024, 41(07): 148-160.

- [13] Niu Xiaojian, Qiang Haofan, Lv Bin, et al. Corporate Cross-border Competition and Debt Default Risk: Machine Learning and Complex-Network-Based Empirical Research [J]. Journal of Financial Research, 2024, (03): 150-168.
- [14] Ding Yue, Fang Kuangnan, Lan Wei. Categorical-Variable Prediction Research Based on Network Relationships [J]. Statistical Research, 2024, 41(01): 148-156.
- [15] Yang Zihui, Zhang Pingmiao, Lin Shihan. Research on Risk Spillover and Prediction across Stock and Bond Markets: Frontier Perspectives from Machine Learning [J]. Journal of Financial Research, 2024, (01): 131-149.
- [16] Wang Lei, Li Xiaoteng, Zhang Zili. Does Credit-Risk Contagion Affect Bond Pricing? Empirical Research Based on Guarantee-Network Big Data [J]. Journal of Financial Research, 2022, (07): 171-189.
- [17] Zhang Liying, Yang Ruojin. Research on Application of Machine-Learning-Based Personal-Loan-Default-Prediction Models [J]. Journal of Financial Regulation Research, 2022, (06): 46-59.
- [18] Cheng Tianxiao, Wen Yuechun. Research on Personal-Customer Default-Probability Measurement for Margin Trading [J]. Journal of Financial Research, 2016, (04): 174-189.