Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 4, No. 1, 2026

Behavioral Sequence Modeling for Early Identification of Payment Default Risk

Alessandro Ferraro [1], Chiara Moretti [2, *], Matteo Rinaldi [1]

[1] Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

[2] Department of Information Engineering, University of Padua, Padua, Italy

Abstract: Card issuers need to recognize clients who are drifting toward payment default early enough to intervene. This paper studies behavioral sequence modeling for this purpose on the public Default of Credit Card Clients data set, in which each of 30,000 clients is described by six monthly records of repayment status, billed amount and paid amount. Early identification is formulated as default prediction from truncated observation windows, so that the lead time between the last observed month and the outcome month ranges from one to six months. We develop SATA-GRU, a gated recurrent encoder over monthly behavioral tokens with static-aware temporal attention and profile fusion, and a hybrid early-warning score that averages SATA-GRU with gradient-boosted trees. Under five stratified splits and against eight baselines, SATA-GRU attains the highest AUC (0.7831), average precision and F1 score and the lowest Brier score among single models on the full window, and the hybrid score raises the AUC to 0.7852, exceeding LightGBM on every split. The hybrid score is also the most accurate method at every lead time, with AUC rising from 0.6874 six months ahead to 0.7852 one month ahead, and it ranks best among clients who default without any observed delinquency. Ablation attributes the gains mainly to derived repayment ratios, profile fusion and status embeddings, and the attention weights concentrate on the earliest observed months of persistently delinquent clients.

Keywords: Credit risk, payment default, behavioral scoring, sequence modeling, gated recurrent unit, temporal attention, early warning.

1. Introduction

Consumer credit portfolios are exposed to losses that materialize months after the first signs of financial distress appear in repayment records. Credit scoring, the statistical assessment of whether a borrower will repay, has therefore become a core decision tool for lenders [1–3]. While application scoring evaluates applicants at the time of the credit decision, behavioral scoring updates the risk assessment of existing accounts with their observed repayment performance [2, 4, 5]. For revolving products such as credit cards, the value of a behavioral score depends on its timing: an account flagged several months before default leaves room for credit-line reduction, payment plans or collection actions, whereas a warning issued in the month of default has little practical value [6, 7].

Most credit scoring studies represent repayment histories as fixed-length feature vectors, either by flattening monthly records or by aggregating them into summary statistics, and then apply logistic regression, neural networks or tree ensembles [8–10]. On such tabular inputs, gradient-boosted decision trees are difficult to outperform [11–13]. Repayment behavior, however, is an ordered process: the month in which a delay first appears, whether balances keep rising and how much of each statement is repaid are trajectory properties that flattened vectors expose only implicitly [14, 15]. Dynamic and survival formulations address the timing of default [16–18], yet they usually depend on manually designed time-varying covariates.

Sequence models offer an alternative. Recurrent and attention-based encoders learn representations directly from event sequences and are the standard tool for user behavior modeling in recommender systems [19–21] and for clinical risk prediction from patient histories [22]. In credit risk, convolutional and recurrent networks have been applied to account balance histories and transaction streams [23, 24], and transformers to corporate default term structures [25]. Two questions remain open for monthly card-account data: how early behavioral sequences can identify future defaulters, and whether a sequence encoder adds information beyond a strong tree ensemble at each lead time. The first question is a form of early classification of time series, in which accuracy is traded against earliness [26, 27].

This paper addresses both questions on the widely used Default of Credit Card Clients data set [8, 28]. The contributions are threefold. First, early identification is formulated as default prediction from truncated observation windows, which yields a controlled lead time of one to six months between the last observed month and the outcome month. Second, SATA-GRU is developed: monthly repayment status, billing and payment records are encoded as behavioral tokens, processed by a gated recurrent unit, pooled by a temporal attention whose scores are conditioned on the client profile, and fused with profile features; a hybrid early-warning score further averages SATA-GRU with LightGBM. Third, a systematic evaluation against eight baselines over five stratified splits covers six lead times, the subgroup of clients who default without any observed delinquency, component ablation and attention analysis.

The remainder of the paper is organized as follows. Section II reviews related work, Section III describes the method, Section IV presents the experimental setup, Section V reports the results, and Section VI concludes the paper.

2. Related Work

2.1. Credit and Behavioral Scoring

Statistical credit scoring has a long tradition, reviewed by Hand and Henley [1], Thomas [2] and Abdou and Pointon [29]. Large benchmarks compare classifiers across portfolios

and show that ensembles often lead while logistic regression remains a robust reference [9, 10, 30, 31]; class imbalance is a recurrent difficulty [32, 33]. Behavioral scoring uses the performance of existing accounts [4, 34], and Kennedy et al. [5] showed that the choice of performance and outcome windows strongly influences behavioral scorecards, which motivates the explicit lead-time analysis in this work. The timing of default has been modeled with survival analysis and proportional hazards [16, 17, 35], extended to macroeconomic covariates [36] and mixture cure models [37], while Markov transition models describe migrations between delinquency states [15]. Dynamic models have been used for forecasting and stress testing card default [18], and Till and Hand [14] modeled card usage behavior directly.

2.2. Machine Learning for Default Prediction

Neural networks [38], support vector machines [39] and tree ensembles [40, 41] have all been applied to credit scoring. On the Taiwanese card data used here, Yeh and Lien [8] compared six data mining techniques for estimating the probability of default, and Hamori et al. [42] found boosting to outperform bagging, random forests and deep neural networks. Khandani et al. [6] and Butaru et al. [7] demonstrated on large proprietary portfolios that machine-learning forecasts of card delinquency translate into substantial loss savings. Whether deep learning helps on tabular credit data is debated: Gunnarsson et al. [11] reported that XGBoost [43] outperforms deep networks, in line with general tabular benchmarks [12, 13]. Deep models are more successful when the input has a sequential structure, such as daily account balances [23], card transactions [24] or multi-channel panel data [25].

2.3. Sequence Modeling and Early Classification

Gated recurrent units such as the long short-term memory [44] and the gated recurrent unit (GRU) [45] capture temporal dependencies while mitigating vanishing gradients [46]. Attention mechanisms [47] and transformers [48] weigh sequence elements adaptively, and temporal convolutional networks are competitive sequence models [49]. Deep learning more broadly [50] has produced encoder–decoder models for sequences [51] and strong time series classifiers [52]. In behavior modeling, attention over user histories underlies deep interest networks [19, 20, 53] and self-attentive recommenders [21], and reverse-time attention provides interpretable risk predictions from clinical visit sequences [22]. Early classification of time series studies predictions from incomplete sequences and the associated earliness–accuracy trade-off [26, 27]. The present work brings these ideas to monthly card-account behavior and quantifies the trade-off for default prediction.

3. Methodology

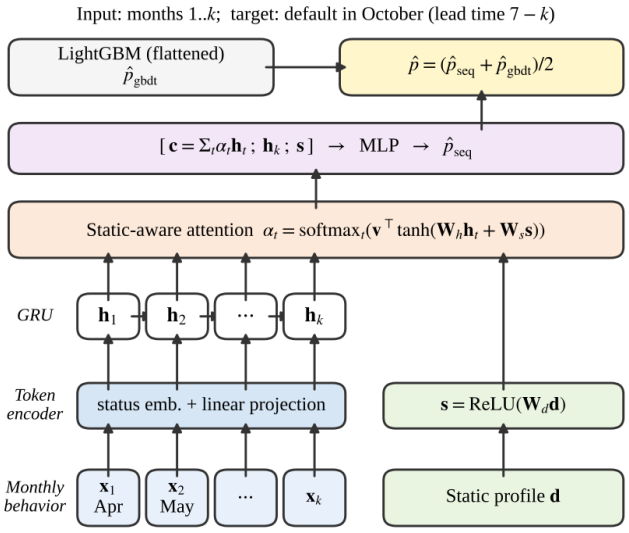

Figure 1. Framework of SATA-GRU and the hybrid early-warning score. Monthly behavioral tokens of the first k months are encoded by a GRU, pooled by static-aware temporal attention, fused with the profile embedding, and averaged with a LightGBM score.

3.1. Problem Formulation

Let client i be observed over T = 6 billing months t = 1, …, T (April to September 2005). In month t the record contains the repayment status $q_{i,t} \in \{-2, \ldots, 8\}$, the billed amount $B_{i,t}$ and the paid amount $P_{i,t}$; the client further has a credit limit L_i and a static profile vector d_i (age, sex, education and marital status). The target $y_i \in \{0, 1\}$ indicates default on the payment due in October 2005. For early identification, a model f_k that observes only the first k months issues the score $\hat{p}_i = f_k(X_{i,1:k}, d_i)$ at a lead time of $\ell = T + 1 - k$ months before the outcome month. Setting k = 6 recovers the conventional full-window task, and k = 1 corresponds to a six-month early warning.

3.2. Behavioral Tokens

Each month is converted into a behavioral token that consists of the categorical status qt and the numeric vector

$$x_t = [\text{slog } B_t, \log(1+P_t), u_t, \rho_t, q_t, \text{slog}(B_t - B_{t-1})] \quad (1)$$

Where $\text{slog}(z) = \text{sign}(z)\log(1+|z|)$ handles negative balances. The utilization and the repayment coverage are

$$u_t = \text{clip}(B_t/L, -0.5, 2), \ \rho_t = \text{clip}(P_t/B_{t-1}, 0, 2) \quad (2)$$

The coverage relates each payment to the preceding statement because the payment of month t settles the bill issued in month t−1; it is set to zero when $B_{t-1} \le 0$ and for the first month, whose balance change is also set to zero. Numeric features are standardized with training-set statistics. Table I summarizes the variables.

3.3. SATA-GRU Encoder

Token and sequence encoding. The status is embedded into an 8-dimensional vector $E(q_t)$ and projected together with x_t; a GRU [45] with 64 hidden units then processes the token sequence:

$$e_t = \text{ReLU}(W_e[x_t; E(q_t)] + b_e), \ h_t = \text{GRU}(e_t, h_{t-1}) \quad (3)$$

The reset and update gates of the GRU decide how much of the previous state is carried forward, which lets the encoder retain early delinquency episodes while tracking recent balance dynamics. The static profile is encoded as

$$s = \mathrm{ReLU}(W_d\, d + b_d) \quad (4)$$

Static-aware temporal attention. The relevance of a month for future default depends both on what happened in that month and on who the client is; a delay carries different information for a young client with a low limit than for an established client with a high limit. The attention score of month t is therefore conditioned on the profile embedding,

$$a_t = v^\top \tanh(W_h\, h_t + W_s\, s + b_a),\ \alpha_t = \exp(a_t)/\textstyle\sum_{j=1}^{k} \exp(a_j) \quad (5)$$

And the context vector $c = \sum_{t=1}^{k} \alpha_t\, h_t$ summarizes the trajectory. Setting $W_s = 0$ recovers profile-agnostic additive attention [47].

Fusion and output. The context vector, the final state and the profile embedding are concatenated and passed through a two-layer perceptron,

$$\hat{p}_{seq} = \sigma(w_2^\top \mathrm{ReLU}(W_1[c;\, h_k;\, s] + b_1) + b_2) \quad (6)$$

With dropout [54] of 0.2 on the hidden layer. Because h_k is fed to the output layer directly, the attention pool is free to focus on information that the final state does not already carry. The network is trained by minimizing the binary cross-entropy

$$\mathscr{L} = -1/N \textstyle\sum_{i=1}^{N} [y_i \log\hat{p}_i + (1 - y_i)\log(1 - \hat{p}_i)] \quad (7)$$

With the AdamW optimizer [55, 56] and early stopping on the validation AUC.

3.4. Hybrid Early-Warning Score

Tree ensembles split on thresholds of individual monthly values, whereas the recurrent encoder summarizes the ordered trajectory. The two inductive biases are complementary, and ensembles of deep models and boosted trees have been reported to outperform boosted trees alone on tabular benchmarks [13]. The hybrid score averages the probabilities of SATA-GRU and of a LightGBM model [57] trained on the flattened tokens of the same window,

$$\hat{p} = (\hat{p}_{seq} + \hat{p}_{gbdt})/2 \quad (8)$$

Equal weights avoid fitting an additional parameter on the validation data, and since both components are trained with the log-loss, the averaged score remains a probability estimate.

4. Experimental Setup

4.1. Data Set

The Default of Credit Card Clients data set [28] contains 30,000 card holders of a Taiwanese bank with 23 explanatory variables and a binary default indicator [8]; 6,636 clients (22.12%) defaulted. Status codes −2, −1 and 0 denote no consumption, payment in full and use of revolving credit, and codes 1–8 denote payment delays of one to eight months. Education categories 0, 5 and 6 and marital category 0 are merged into "others". Fig. 2 contrasts the monthly trajectories of defaulters and non-defaulters. The share of delinquent defaulters grows from 24.3% in April to 51.7% in September, whereas for non-defaulters it rises only from 6.3% to 14.5%; overall, 64.8% of defaulters but only 24.7% of non-defaulters are delinquent at least once in the window. Defaulters also use a larger share of their limit (48.9% versus 40.4% in September) and repay a smaller fraction of the previous statement (about 20% versus 31%). The divergence accumulates toward the outcome month, so early identification must detect a weaker version of this signal from shorter histories.

Table I. Variables of the Behavioral Sequences

Group	Variables	Description
Profile d, L	LIMIT_BAL, SEX, EDUCATION, MARRIAGE, AGE	credit limit (log), demographics (one-hot)
Status q_t	PAY_0, PAY_2–PAY_6	−2/−1/0: no use/paid/revolving; 1–8: months of delay
Amounts	BILL_AMT1–6, PAY_AMT1–6	billed amount B_t, paid amount P_t (NT$)
Derived	u_t, ρ_t, ΔB_t	utilization, repayment coverage, balance change
Target y	default payment next month	default in Oct. 2005 (22.12% positive)

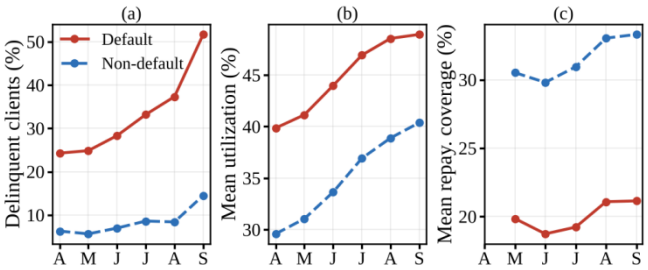

Figure 2. Monthly behavior of defaulters and non-defaulters from April to September 2005: (a) share of clients with a payment delay, (b) mean credit utilization, and (c) mean repayment coverage of the previous statement.

4.2. Baselines

Eight baselines receive exactly the same information as SATA-GRU. Logistic regression (LR) uses standardized features and one-hot status codes; random forest (RF) [40] and LightGBM [57], an efficient implementation of the gradient boosting machine [41], use the flattened raw tokens; the multilayer perceptron (MLP) uses the same inputs as LR. The sequence baselines share the token encoder and the profile branch of SATA-GRU and differ in the encoder. LSTM [44] and GRU [45] use the final hidden state, the temporal convolutional network (TCN) [49] applies two causal convolutions with dilations 1 and 2, and the Transformer [48] uses two pre-norm self-attention layers with four heads, learnable month embeddings and mean pooling.

4.3. Protocol and Metrics

The data are split five times into stratified training, validation and test sets of 70%, 10% and 20% (21,000, 3,000 and 6,000 clients) with different random seeds; every model is trained on every split and the results are averaged. Ranking quality is measured by the area under the ROC curve (AUC) [58], by the average precision (AP), which is more informative under class imbalance [33, 59], and by the Kolmogorov–Smirnov statistic $KS = \max_\tau[TPR(\tau) - FPR(\tau)]$ commonly used in credit scoring. The F1 score is computed at the threshold that maximizes F1 on the validation set, the Brier score measures the accuracy of the predicted probabilities, and R@20 is the share of defaulters captured among the 20% of clients with the highest scores, which corresponds to a fixed review capacity of an early-warning process.

The neural models are implemented in JAX with a token dimension of 32, 64 hidden units, a 16-dimensional profile embedding and a 32-dimensional attention space. They are trained with a learning rate of 10^{-3}, weight decay of 10^{-3} and batch size 256 for at most 40 epochs, with early stopping after

six epochs without improvement of the validation AUC. LightGBM uses a learning rate of 0.03, 31 leaves, at least 50 samples per leaf, row and column subsampling of 0.8 and early stopping after 100 rounds; RF uses 300 trees with at least 10 samples per leaf, and LR and RF are implemented with scikit-learn [60]. Hyperparameters were chosen on validation data and kept fixed for all splits and windows.

5. Results and Discussion

5.1. Overall Comparison

Table II. Default Prediction with the Full Window (k = 6)

Model	AUC	AP	KS	F1	Brier	R@20
LR	77.49±0.75	54.10	41.83	52.81	13.64	50.20
RF	78.01±0.82	55.12	42.97	54.05	13.52	50.87
LightGBM	78.24±0.70	55.38	43.26	53.90	13.75	51.27
MLP	78.06±0.52	54.99	42.83	53.85	13.52	50.79
LSTM	78.26±0.60	55.27	42.87	53.75	13.49	51.00
GRU	78.24±0.59	55.17	43.02	53.92	13.50	51.02
TCN	78.04±0.55	55.06	42.80	53.61	13.50	51.02
Transformer	77.92±0.71	54.96	42.82	53.80	13.56	50.50
SATA-GRU	78.31±0.67	55.48	43.17	54.14	13.48	51.05
Hybrid	78.52±0.70	55.83	43.23	53.99	13.48	51.21

All metrics ×100; mean over five stratified splits, AUC also with standard deviation. Lower Brier is better. Hybrid = SATA-GRU + LightGBM. Best values in bold.

Table II reports the results with the full window. All models lie within a narrow band of 77.5–78.5 AUC points, consistent with earlier studies on this data set [8], [42] and with the strength of tree ensembles on tabular data [12]. Among single models, SATA-GRU obtains the highest AUC (78.31), AP (55.48) and F1 (54.14) and the lowest Brier score (13.48), while LightGBM achieves the highest KS (43.26) and R@20 (51.27). The margins of SATA-GRU over GRU (+0.07 AUC points, better in four of five splits) and LightGBM (+0.07) are small, whereas its improvements over the Transformer (+0.39, five of five splits, paired t-test $p = 0.021$) and LR (+0.82, $p = 0.002$) are significant. The recurrent encoders outperform the Transformer and the TCN, indicating that with six monthly steps the inductive bias of a recurrence suits the data better than global self-attention or convolution. The hybrid score achieves the best AUC (78.52) and AP (55.83) of all methods and exceeds both LightGBM (+0.28, five of five splits, $p = 0.053$) and SATA-GRU (+0.21, five of five splits, $p = 0.019$).

Fig. 3 plots the ROC and precision–recall curves of the first split. On a single CPU core, SATA-GRU trains in about 12 s per split, compared with 1.3 s for LightGBM and 34 s for the Transformer, so the hybrid score adds little computational cost.

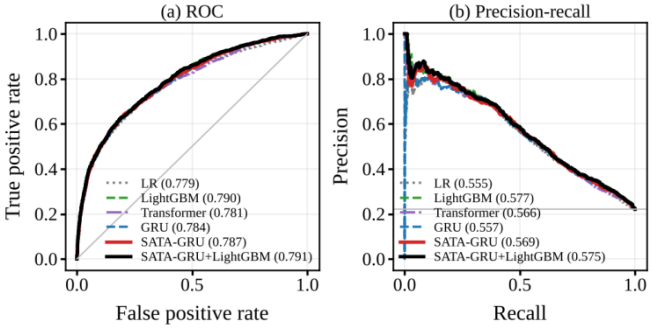

Figure 3. ROC (a) and precision–recall (b) curves on the test set of the first split with the full window; AUC and AP are given in parentheses.

5.2. Early Identification

Table III. AUC (×100) for Observation Windows of k Months (Lead Time 7 − k)

k (lead time)	1 (6)	2 (5)	3 (4)	4 (3)	5 (2)	6 (1)
LR	68.12	69.53	71.00	72.55	74.56	77.49
LightGBM	68.50	70.17	71.44	73.25	75.29	78.24
MLP	68.24	69.79	71.20	72.95	75.05	78.06
GRU	68.46	69.71	71.35	73.18	75.31	78.24
SATA-GRU	68.36	69.92	71.52	73.20	75.25	78.31
Hybrid	68.74	70.29	71.74	73.44	75.51	78.52
Clients without observed delinquency						
LR	62.89	63.51	64.00	64.64	65.34	64.10
LightGBM	63.35	64.51	64.69	65.89	66.78	66.23
MLP	63.00	63.85	64.23	65.23	66.31	65.71
GRU	63.28	63.76	64.40	65.67	66.74	66.28
SATA-GRU	63.16	64.11	64.67	65.76	66.57	66.38
Hybrid	63.67	64.64	65.04	66.16	67.06	66.78

Mean over five splits; best value per column in bold. The lower block evaluates clients without any payment delay in the observed window.

Table III and Fig. 4 report the performance as a function of the observation window. Accuracy decreases smoothly as the lead time grows: with five months of history the hybrid AUC is 75.51, and even a single month observed six months before the outcome yields 68.74. The hybrid score is the most accurate method at every lead time and improves over LightGBM in all five splits for $k = 3, 4, 5$ and 6 ($p \le 0.053$).

SATA-GRU alone exceeds GRU for k = 2, 3, 4 and 6 and stays within 0.1 AUC points of LightGBM for $k \geq 3$, whereas LightGBM leads the single models for $k \leq 2$, where a sequence of one or two steps offers little temporal structure. In terms of R@20 in Fig. 4(b), flagging the 20% riskiest clients with the hybrid score identifies 37.9% of future defaulters six months ahead, 46.9% two months ahead and 51.2% one month ahead, against 20% for random selection.

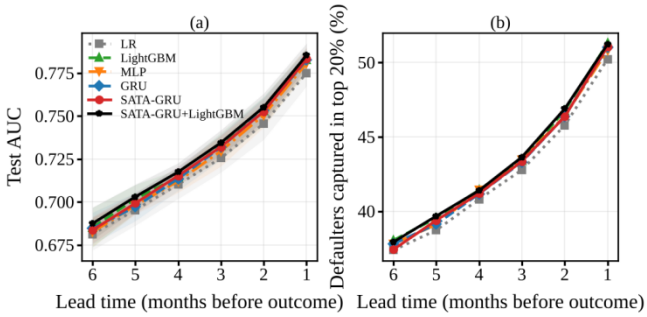

Figure 4. Early identification performance versus lead time (mean over five splits, shaded band ±1 std): (a) test AUC and (b) share of defaulters captured among the 20% highest-scored clients.

5.3. Clients without Observed Delinquency

Defaulters who never miss a payment within the observed window are the hardest to anticipate and the most valuable to detect early. Their share is substantial: with k = 6, 36% of the test defaulters (478 of 1,327 on average) show no delay in any of the six months. The lower block of Table III reports the AUC within this subgroup, which lies 5 to 13 points below the overall AUC because the strongest signal, an observed delay, is absent. The hybrid score again ranks best at every window (63.67–67.06), followed by LightGBM and the recurrent models, and its advantage over LR reaches 2.7 points at k = 6. The nonlinear models therefore extract risk cues from balances, utilization and repayment coverage that go beyond delinquency codes.

5.4. Ablation Study

Table IV. Ablation of SATA-GRU (×100, Mean over Five Splits)

	k = 3		k = 6		
Variant	AUC	KS	AUC	AP	KS
SATA-GRU (full)	71.52	31.68	78.31	55.48	43.17
w/o static conditioning	71.55	31.64	78.32	55.49	43.12
w/o temporal attention	71.35	30.93	78.30	55.56	43.12
w/o profile branch	71.23	30.84	78.11	55.19	42.97
w/o status embedding	71.42	31.53	78.07	54.69	43.06
w/o derived features	71.09	30.86	77.90	55.15	43.14

Static conditioning: $W_s = 0$ in (5). Derived features: u_t, ρ_t and the balance change.

Table IV removes one component at a time. Dropping the derived features causes the largest AUC loss at both windows (−0.43 at k = 3 and −0.41 at k = 6), confirming that utilization and repayment coverage are informative beyond the raw amounts. Removing the profile branch costs 0.29 and 0.20 AUC points, and replacing the status embedding by the numeric code reduces AUC by 0.24 and AP by 0.79 points at k = 6. Temporal attention matters for the short window: without it, KS at k = 3 falls from 31.68 to 30.93, while at k = 6 the final state already summarizes the trajectory and both variants perform alike. Conditioning the attention scores on the profile leaves accuracy unchanged (differences of at most 0.05 points), which shows that the profile information is exploited through the fusion layer; the conditioning makes the attention distribution profile-specific, which supports the analysis below.

5.5. Attention Analysis

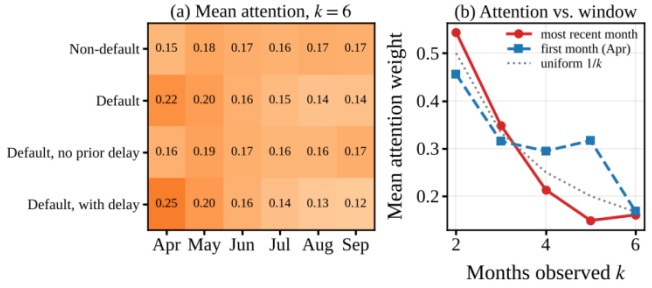

Figure 5. Attention weights of SATA-GRU on the test set of the first split: (a) mean weight per month and client group for k = 6; (b) mean weight of the first and the most recent observed month for different window lengths.

Fig. 5(a) averages the attention weights over test clients. For non-defaulters the weights are close to uniform (0.15–0.18), whereas for defaulters with at least one delay they decrease from 0.25 in April to 0.12 in September. Because h_k enters the output layer directly, the attention context complements it with the early part of the trajectory and thereby encodes how long delinquency has persisted, in line with Markov-type behavioral models in which the delinquency state and its persistence drive future transitions [4, 15]. For defaulters without any delay the weights remain nearly uniform, consistent with the absence of a distinct delinquency onset. Fig. 5(b) shows that the most recent month receives the largest weight for $k \leq 3$, the first month dominates for k = 4 and 5 (0.29 and 0.32), and at k = 6 the population-average weights are almost uniform while the differentiation appears between client groups.

6. Conclusion

This paper investigated behavioral sequence modeling for the early identification of payment default. Formulating the task as prediction from truncated monthly windows quantifies the accuracy available at each lead time: on the Taiwanese credit card data, the hybrid score reaches an AUC of 0.687 six months ahead and 0.785 one month ahead and is the most accurate method at every lead time. SATA-GRU is the strongest single model on the full window and contributes information that complements gradient-boosted trees, which makes the hybrid score consistently better than LightGBM. Derived repayment ratios, profile fusion and status embeddings are the most important components, and the attention weights reveal how the persistence of delinquency shapes the risk assessment. Future work will extend the evaluation to longer histories and transaction-level data, add multi-horizon outputs in the spirit of survival models, and study the stability of early-warning scores under portfolio shifts.

References

[1] Xing, S., Wang, Y., & Liu, W. (2025). Multi-dimensional anomaly detection and fault localization in microservice architectures: A dual-channel deep learning approach with causal inference for intelligent sensing. Sensors, 25(11), 3396.

[2] Chen, J., & Fan, H. (2025). Beyond automation in tax compliance through artificial intelligence and professional judgment. Frontiers in Business and Finance, 2(2), 399–418.

[3] Xing, S., Wang, Y., & Liu, W. (2025). Self-adapting CPU scheduling for mixed database workloads via hierarchical deep reinforcement learning. Symmetry, 17(7), 1109.

[4] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement learning with thinking model for dynamic supply chain network design. IEEE Access, 12, 195974–195985.

[5] Wang, Y., & Zhang, C. (2023). Aeris: Adaptive executor-scoped resource management for privacy-preserving data processing in Apache Spark. Eurasia Journal of Science and Technology, 5(2), 25–33.

[6] Chen, X. (2024). Who is accountable for AI decisions? Rethinking legal liability and ethical responsibility through counterfactual harm apportionment. Innovation and Technology Studies.

[7] He, X., & Zhou, L. (2024). Adaheal: An adaptive framework for fault detection and automated recovery in large-scale cloud-native microservices. AI and Data Science Journal, 1(1), 132–141.

[8] Wang, J., Ge, Y., Liu, J., & Jiang, B. (2024). Early warning of liquidity stress in U.S. banks with seasonally aligned frailty hazards and delayed-feedback conformal alarms. Journal of Trends in Finance and Economics, 1(1), 85–100.

[9] Qin, Y., Rhee, M., Teng, D., Zhang, C., & Zi, B. (2026). Neuro-symbolic constraint verification for LLM-driven internet finance transaction execution. Scientific Reports.

[10] Fan, H., Jiao, Y., Wang, M., Chen, J., & Yang, S. (2026). Continual graph learning under strategic adversarial drift in dynamic networks. Scientific Reports.

[11] Zhou, F. (2026). Spectral co-design of ship rudder geometry and heading autopilot for energy efficiency and manoeuvring safety. Transactions on Environment, Energy and Earth Sciences, 7, 158–181.

[12] Li, P., Feng, E., Wang, G., & Mo, T. (2024). Hybrid deep reinforcement learning and integer optimization for resource scheduling in heterogeneous AI computing clusters. Innovation and Technology Studies, 1(1), 48–56.

[13] Li, P., Bian, M., & Lin, Y. (2026). Carbon-aware offline reinforcement learning for joint inventory replenishment and transportation scheduling in multi-echelon supply chains. Highlights in Business, Economics and Management, 69, 251–262.

[14] Zhang, F., & Yang, J. (2022). Accuracy-aware model compression for real-time visual perception under resource constraints. Journal of Computer Science and Electrical Engineering, 4(1), 45–52.

[15] Jiao, Y., & Liang, Y. (2024). An integrated process-mining and explainable machine-learning framework for bottleneck and internal-control risk detection in the multi-entity financial close. Social Science and Management, 1(3), 96–107.

[16] Zhou, F., & Liu, W. (2026). Physics-anchored residual learning for ship fuel consumption and carbon dioxide emission prediction. Energy and AI, 100945.

[17] Wang, J., Liu, J., Zheng, W., & Ge, Y. (2025). Temporal heterogeneous graph contrastive learning for fraud detection in credit card transactions. IEEE Access.

[18] Shen, A., & Huang, S. (2022). A systems-aware framework for intelligent large-scale recommendation models. Journal of Computer Science and Electrical Engineering, 4(1), 36–44.

[19] Lu, Z., Zou, X., & Wei, K. (2022). Machine learning-based detection of BGP routing anomalies for enhanced network security and resilience: A hybrid multi-scale temporal learning approach. Journal of Computer Science and Electrical Engineering, 4(1), 62–71.

[20] Hu, X., Guo, L., Wang, J., & Liu, Y. (2025). Computational fluid dynamics and machine learning integration for evaluating solar thermal collector efficiency-based parameter analysis. Scientific Reports, 15, 24528.

[21] Xing, S., & Wang, Y. (2025). Proactive data placement in heterogeneous storage systems via predictive multi-objective reinforcement learning. IEEE Access, 13, 117986–117998.

[22] Xing, S., & Wang, Y. (2022). MCG: Marginal–conditional graph learning for joint anomaly detection and root cause localization in distributed cloud systems. Journal of Computer Science and Electrical Engineering, 4(1), 81–88.

[23] Wang, Y., Bian, M., & Lin, Y. (2022). An explainable temporal graph neural network for disruption prediction and risk propagation in multi-tier supply chains. Journal of Computer Science and Electrical Engineering, 4(1), 17–25.

[24] Xu, Q. (2023). Explainable machine learning for tax reporting anomaly detection and compliance risk assessment. World Journal of Economics and Business Research, 1(1), 56–66.

[25] Zou, J., Qin, Y., & Rhee, M. (2024). Beyond model scale: A task-aware reliability benchmark for large language models in clinical decision support. Innovation and Technology Studies, 1(1), 39–47.

[26] Fan, H., & Chen, J. (2025). From ledgers to large language models in contemporary accounting practice. American Journal of Big Data, 6(5).

[27] Ping, W., & Fan, H. (2024). Disclosure quality and valuation risk in Level 3 investments: A disclosure-implied risk framework for U.S. business development companies. Social Science and Management, 1(1), 78–87.

[28] Wang, J., Liu, J., Ge, Y., & Jiang, B. (2023). Risk-aware ensemble learning for credit default prediction under economic uncertainty. World Journal of Economics and Business Research, 1(1), 18–32.

[29] Ji, E., Wang, Y., Xing, S., & Jin, J. (2025). Hierarchical reinforcement learning for energy-efficient API traffic optimization in large-scale advertising systems. IEEE Access.

[30] Qin, Y., Zou, J., & Rhee, M. (2022). An AI-driven intelligent transaction execution framework for internet finance platforms: Cost-sensitive real-time integrity interdiction. Eurasia Journal of Science and Technology, 4(2), 36–44.

[31] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). Sensor-integrated inverse design of sustainable food packaging materials via generative adversarial networks. Sensors, 25(11), 3320.

[32] Mo, T., Zhang, C., Zou, J., Guo, Z., & Rhee, M. (2026). Self-evolving AI agents with dual memory for automated software testing and bug localization. IEEE Access.

[33] Zou, X., Wei, K., & Lu, Z. (2024). AEGIS: Calibrated anomaly-risk-driven checkpointing and proactive migration for fault-tolerant high-concurrency stream processing. Innovation and Technology Studies.

[34] Zhang, F., & Wu, B. (2025). Large language models as general purpose intelligence systems for reasoning, planning and decision making. American Journal of Artificial Intelligence and Neural Networks, 6(4), 45–72.

[35] Zhang, F., & Yang, J. (2024). Robustness-aware model compression for multimodal perception under challenging conditions. AI and Data Science Journal, 1(1), 103–111.

[36] Wang, G., Li, P., Feng, E., & Mo, T. (2024). Uncertainty-aware LLM-as-a-judge evaluation for reliable search and recommendation systems: Calibrated confidence and impact-

aware selective human review. AI and Data Science Journal, 1(1), 112–121.

[37] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.

[38] Wei, K., Lu, Z., & Zou, X. (2024). A scalable causal machine learning framework for data-driven decision optimization. AI and Data Science Journal, 1(1), 122–131.

[39] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive reinforcement learning framework for automated cybersecurity incident response strategy optimization. Applied Sciences, 15(2), 951.

[40] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 12, 116546–116557.

[41] Chen, X. (2023). Algorithmic bias and discrimination in AI-driven lending: A Rashomon-based audit of less discriminatory alternatives and adverse-action accountability. World Journal of Sociology and Law.

[42] Ping, W., & Fan, H. (2023). Beyond isolated ratios: Accrual–revenue–cash flow indicators and interaction-aware learning for detecting financial reporting irregularities in U.S. public companies. World Journal of Economics and Business Research, 1(1), 46–55.

[43] Chen, X. (2026). Can people challenge an AI decision? Measuring the contestability chain in law and corporate disclosure. AI and Ethics, 6(6), 642.

[44] Rhee, M., Zou, J., & Qin, Y. (2024). CLARITY: A closed loop framework for duplicate aware failure prediction and budgeted automated remediation in cloud infrastructure. AI and Data Science Journal, 1(1), 76–85.

[45] Lin, Y., Yue, X., Bian, M., Zhu, R., & Teng, D. (2026). Subgraph-level temporal graph transformer with laundering-motif prototypes for anti-money laundering on blockchains. IEEE Access.

[46] Zhou, L., & He, X. (2022). Dependency-aware root cause analysis and automated fault recovery for cloud-native microservices. Journal of Computer Science and Electrical Engineering, 4(1), 72–80.

[47] Chen, J., Zhang, X., & Han, X. (2023). Explainable identification of tax-related financial reporting anomalies in U.S. public companies: Cross-fitted tax-consistency residuals and Owen-value Shapley propagation. World Journal of Economics and Business Research, 1(1), 33–45.

[48] Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. Applied Sciences, 14(20), 9213.

[49] Yao, W. (2024). PrivTok-XR: Privacy-preserving real-time speech processing for accessible extended reality systems. Innovation and Technology Studies.

[50] Wang, J., Tan, Y., Jiang, B., Wu, B., & Liu, W. (2025). Dynamic marketing uplift modeling: A symmetry-preserving framework integrating causal forests with deep reinforcement learning for personalized intervention strategies. Symmetry, 17(4), 610.

[51] Wang, G., Li, P., Feng, E., Shen, A., & Rhee, M. (2026). A dual-process memory-augmented large language model agent for explainable sequential recommendation. IEEE Access.

[52] Zhang, F., & Yang, J. S. (2025). Learning driven decision intelligence for autonomous driving through multimodal understanding world modeling and policy optimization. Frontiers in Artificial Intelligence Research, 2(3), 616–634.

[53] Chen, J., Liang, Y., Liu, J., & Zhou, M. (2026). Temporal transformer with conditional tabular GAN for credit card fraud detection: A sequential deep learning approach. Mathematics, 14(7), 1183.

[54] Mo, T., Wang, G., Feng, E., & Li, P. (2023). Execution-feedback-guided context selection for large language model-based software testing and debugging. World Journal of Information and Knowledge Management, 1(1), 55–64.

[55] Wang, Y., & Xing, S. (2025). AI-driven CPU resource management in cloud operating systems. Journal of Computer and Communications, 13(6), 135–149.

[56] Liang, Y., & Jiao, Y. (2024). Real-time financial fraud detection using a feature-light temporal dynamic graph neural network with contrastive learning. Journal of Trends in Finance and Economics, 1(1), 76–84.

[57] Xu, Q., & Zhu, R. (2024). CoSTAR: Explainable and cost-sensitive machine learning for tax-aware corporate fraud risk detection. Innovation and Technology Studies.

[58] Wang, B., Wang, Z., Zhao, W., & Liu, Y. (2025). Network fabric simulation and validation for data center routing convergence under large-scale failure scenarios. Computer Science Bulletin, 8(1), 310–326.

[59] Zhou, F. (2026). Moment-constrained side-force allocation for wind-assisted merchant ships: Hydrodynamic performance of retractable anti-leeway fins and flap rudders. Innovation and Technology Studies, 3(1), 27–41.

[60] Yao, W. (2022). Raptor: Risk-adaptive security and privacy-preserving transaction processing for large-scale mobile payment systems. Journal of Computer Science and Electrical Engineering, 4(1), 89–97.

[61] Lin, Y., Wang, Y., & Bian, M. (2024). XOR-EW: An explainable and cost-aware artificial intelligence framework for operational risk early warning in financial institutions. AI and Data Science Journal, 1(1), 94–102.