Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37042

Intelligent Identification of Exercise Rehabilitation Service Discontinuation Risk Driven by Social Ecological Characteristics

Yuejuan Dai

Xi'an Physical Education University, Xi'an, China
daiyuejuan02539@outlook.com

Abstract. Early discontinuation from community exercise rehabilitation is rarely determined by one clinical factor and is more often produced by interacting individual, interpersonal, organizational, community, and early-participation conditions. This study developed a Social-Ecological Hierarchical Attention Network to identify discontinuation at the end of week two among adults enrolled in twelve-week programs at six community health centers. The analytic cohort comprised 3,684 episodes, of which 1,079 met the prespecified fourteen-day non-attendance outcome. The proposed model preserved five ecological feature groups, represented missingness explicitly, learned cross-level interactions, and produced isotonic-calibrated probabilities. In temporal testing, the model achieved an AUROC of 0.861 ± 0.006, an AUPRC of 0.708 ± 0.009, a Brier score of 0.129 ± 0.003, and a calibration slope of 0.98 ± 0.04. Removing early participation features reduced AUPRC by 0.071 ± 0.008, while direct feature concatenation reduced it by 0.039 ± 0.006. The findings indicate that combining early attendance behavior with organizational accessibility and interpersonal support provides accurate, calibrated, and interpretable identification of rehabilitation service discontinuation risk.

Keywords: exercise rehabilitation, service discontinuation, social ecology, risk identification, hierarchical attention

1. Introduction

Sustained participation is necessary for exercise rehabilitation to produce clinically meaningful gains in mobility, symptom control, cardiovascular capacity, and self-management. In routine services, however, discontinuation may develop before formal withdrawal is documented. Missed supervised sessions, delayed rescheduling, declining home-exercise submission, and reduced responsiveness to staff can accumulate during the first weeks and eventually create a prolonged absence. Recent reviews show that exercise adherence is measured inconsistently and is affected by personal, treatment-related, provider-related, and health-system conditions [1]. In exercise-based physical therapy, logistical access and insufficient authorization can be as influential as diagnosis or symptom response [2]. Clinical studies also indicate that functional status, disease burden, and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37042

exercise capacity shape adherence, but their effects vary substantially across populations and program designs [3].

A purely individual risk model therefore omits mechanisms operating around the participant. Qualitative evidence identifies social support, perceived benefit, scheduling options, and practical attendance barriers as central to continued engagement [4]. Pulmonary rehabilitation research similarly distinguishes program uptake, persistence, completion, and treatment response, showing that these are related but non-equivalent outcomes [5]. Meta-analytic evidence further suggests that supervision and intervention intensity influence dropout beyond baseline clinical characteristics [6]. Such findings support a social-ecological representation in which individual limitations interact with family assistance, therapist continuity, organizational flexibility, transportation, and neighborhood resources.

Existing prediction models usually concatenate heterogeneous variables into one tabular vector. This strategy can capture average associations but obscures which ecological level is generating risk and whether cross-level combinations alter the prediction. The present study develops a Social-Ecological Hierarchical Attention Network that retains five feature groups, models missingness, learns within-level and cross-level dependencies, calibrates individual probabilities, and generates ecological contribution profiles. The prediction point is fixed after fourteen observable days, enabling intervention before sustained disengagement becomes administratively established. Performance is tested temporally against regularized logistic regression, XGBoost, and FT-Transformer using discrimination, calibration, threshold performance, ablation, and subgroup stability.

2. Literature review

Exercise rehabilitation discontinuation is commonly operationalized through incomplete sessions, prolonged absence, failure to submit prescribed activity, or formal withdrawal. The determinants extend beyond symptoms and physical capacity. Social support, self-efficacy, therapist communication, travel burden, scheduling flexibility, cost, and continuity of supervision repeatedly influence whether participants remain engaged. Recent social-ecological synthesis has shown that barriers and facilitators occur simultaneously across intrapersonal, interpersonal, organizational, and environmental levels, with supportive relationships frequently offsetting health limitations [7]. This multilevel structure also clearly explains why the same symptom burden may produce different attendance patterns across clinics and neighborhoods.

Digital prediction studies demonstrate that early behavioral records can identify future service interruption. Appointment no-show models have benefited from explicit missing-data handling, longitudinal attendance history, and individualized explanations [8]. However, rehabilitation data are structurally heterogeneous: symptoms, family assistance, clinic operations, geospatial accessibility, and attendance traces do not have equivalent meanings or measurement processes. Treating them as exchangeable features can weaken interpretability and may conceal interactions such as mobility restriction combined with poor transport availability.

Transformer-based tabular architectures can model nonlinear feature dependencies and often outperform conventional neural baselines, although gradient-boosted trees remain competitive across datasets [9]. Clinical deployment additionally requires more than AUROC. Calibration, probability error, subgroup behavior, and threshold-specific consequences are necessary because an overconfident model may misallocate follow-up resources even when ranking performance is acceptable [10]. A hierarchical architecture is therefore appropriate when prediction must preserve

DOI: 10.54254/2755-2721/2026.37042

ecological organization, support calibrated probabilities, and explain whether risk arises primarily from personal, relational, service, community, or early-participation conditions.

3. Method

3.1. Cohort and outcome construction

Adults aged at least 18 years who entered twelve-week exercise rehabilitation between January 2024 and December 2025 at six community health centers were screened. Each program prescribed two supervised sessions and one documented home session weekly. Episodes were excluded when fewer than seven observable days were available, the participant transferred to inpatient treatment during the first two weeks, or duplicated enrollment records could not be reconciled. The final cohort contained 3,684 episodes from 3,517 individuals; repeated episodes were retained only when separated by six months and remained in the same partition.

The prediction time was 23:59 on day 14. Discontinuation was defined as fourteen consecutive days without supervised attendance before week twelve, provided that no medical suspension, planned completion, relocation, or documented transfer explained the absence. PostgreSQL event windows joined appointments, cancellations, therapist notes, reminder responses, and home-exercise submissions. Two reviewers audited 200 stratified episodes; agreement with the automated definition was 0.94, and discrepancies were adjudicated before modeling. The observed discontinuation rate was 29.3%, with a median event time of 34 days after enrollment.

3.2. Social ecological variables and data processing

The feature space contained 87 variables organized into five matrices. Individual variables represented demographics, diagnostic group, pain, fatigue, mobility limitation, distress, exercise self-efficacy, digital literacy, and prior rehabilitation. Interpersonal variables represented living arrangement, caregiving availability, family support, transport assistance, message responsiveness, and peer participation. Organizational variables included referral-to-start delay, scheduling flexibility, therapist continuity, group size, reminder latency, replacement-session availability, and communication quality. Community variables were derived in QGIS from OpenStreetMap routing and neighborhood layers, including travel time, public-transport availability, walkability, facility density, and area deprivation. Early-participation variables summarized attendance rate, cancellation rate, maximum absence streak, rescheduling latency, home-exercise submission, and deviation from prescribed frequency during days 1–14. Continuous variables were winsorized at the 1st and 99th percentiles and standardized using training-cohort parameters. Bayesian-ridge iterative imputation was fitted within each training fold, while categorical missingness was encoded as an explicit unknown level. A binary missingness mask accompanied every feature so that unrecorded information could contribute differently from an observed neutral value. Variables exceeding 35% missingness were removed before splitting. No post-day-14 information entered the predictors.

3.3. Hierarchical attention model and risk output

Categorical values were mapped to sixteen-dimensional embeddings, numerical values were linearly projected to the same dimension, and missingness embeddings were added before ecological partitioning. For participant (i) and ecological level (l), the feature matrix entered a two-head self-attention encoder with hidden dimension 64. Attention pooling generated one vector per level, after

DOI: 10.54254/2755-2721/2026.37042

which a four-head cross-level block with dimension 128 learned interactions among the five ecological representations as shown in Equation (1)-(3):

$$H_i^{(l)} = LN(X_i^{(l)} + MHA_l(X_i^{(l)}, X_i^{(l)}, X_i^{(l)})) \tag{1}$$

$$z_i^{(l)} = \sum_{j=1}^{m_l} \frac{exp[u_l^\top tanh(W_l h_{ij}^{(l)} + b_l)]}{\sum_{k=1}^{m_l} exp[u_l^\top tanh(W_l h_{ik}^{(l)} + b_l)]} h_{ij}^{(l)} \tag{2}$$

$$\hat{p}_i = \sigma[w_o^\top GELU(W_o MHA(Z_i, Z_i, Z_i) + b_o) + b_o] \tag{3}$$

The representation passed through 128- and 64-unit layers with GELU activation and dropout 0.20. AdamW used a learning rate of 2×10^{-4}, weight decay of 0.01, batch size of 64, and gradient clipping at 1.0. Class imbalance was handled through focal binary cross-entropy. Isotonic regression was fitted only on internal validation predictions. SHAP summarized feature effects, while integrated gradients and ecological attention weights generated individualized level-specific explanations. The complete architecture preserves the five social-ecological feature groups, learns dependencies within and across ecological levels, and produces both calibrated risk probabilities and individualized ecological explanations, as illustrated in Figure 1.

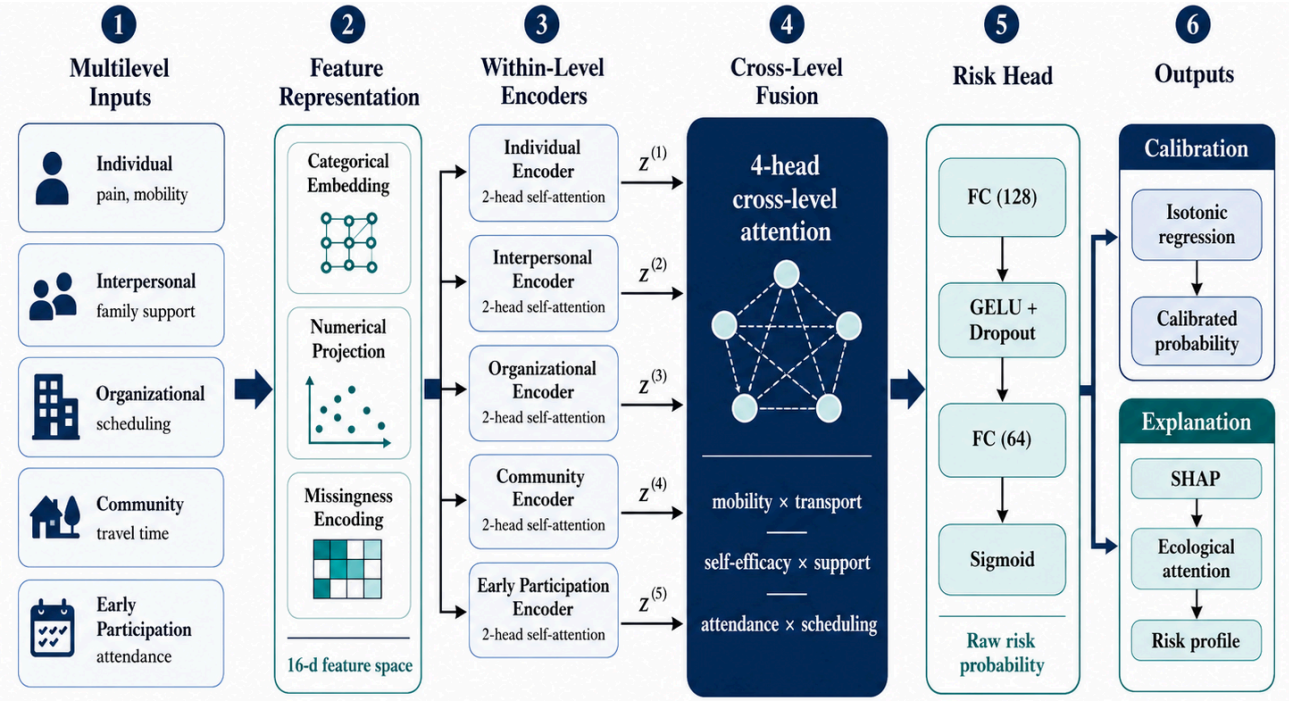

Figure 1. Architecture of the social-ecological hierarchical attention network for rehabilitation discontinuation risk identification

4. Experimental design

4.1. Data partition and model comparison

Enrollments from January 2024 through June 2025 formed the development cohort. Participant-grouped stratification assigned 70% to training and 30% to internal validation. Enrollments from July through December 2025 constituted an untouched temporal test cohort, yielding 1,900 training,

DOI: 10.54254/2755-2721/2026.37042

814 validation, and 970 test episodes. L2-regularized logistic regression, XGBoost, FT-Transformer, and the proposed model received identical predictors, imputations, and outcome labels. Hyperparameters were selected by validation AUPRC without access to temporal data. Models were fitted with seeds 17, 42, and 89. AUROC, AUPRC, F1-score, sensitivity, specificity, Brier score, calibration slope, and expected calibration error were averaged across seeds. The operating threshold was selected on validation data to maintain sensitivity of at least 0.80 while maximizing specificity. Participant-level bootstrap resampling with 1,000 replicates produced confidence intervals and paired two-sided tests against the strongest baseline.

4.2. Ablation and robustness testing

Ablations removed interpersonal, organizational, community, and early-participation matrices. A structural ablation replaced hierarchical encoding with direct concatenation while retaining parameter count within 5%. Changes in AUPRC, Brier score, sensitivity, and calibration error quantified information loss. Robustness analyses reused the fixed validation threshold for participants aged at least 65 years, those living alone, those with mobility limitations, and residents of high-deprivation areas. Leave-one-site-out evaluation tested dependence on local workflow. Explanation stability was calculated as the rank correlation of the ten highest absolute feature contributions across seeds; calibration was considered acceptable when slope remained between 0.90 and 1.10 and expected calibration error was below 0.03.

5. Results

5.1. Risk identification and calibration performance

The proposed model provided the strongest temporal discrimination and lowest probability error. Relative to FT-Transformer, AUROC increased by 0.023 (95% CI: 0.014–0.033, ($p<0.001$)), AUPRC increased by 0.039 (95% CI: 0.026–0.052, ($p<0.001$)), and Brier score decreased by 0.010 (95% CI: 0.006–0.015, ($p=0.002$)) (see Table1). At the locked threshold of 0.318, sensitivity remained above 0.80 while specificity reached 0.752 ± 0.010. The post-calibration slope of 0.98 ± 0.04 and expected calibration error of 0.014 ± 0.003 indicated limited temporal overconfidence.

Table 1. Temporal-test risk identification and calibration performance

Model	AUROC	AUPRC	F1-score	Sensitivity	Specificity	Brier score	Calibration slope	ECE
L2 logistic regression	0.754 ± 0.008	0.506 ± 0.011	0.548 ± 0.010	0.806 ± 0.006	0.552 ± 0.013	0.172 ± 0.004	0.78 ± 0.06	0.061 ± 0.007
XGBoost	0.823 ± 0.006	0.641 ± 0.008	0.638 ± 0.009	0.812 ± 0.009	0.690 ± 0.011	0.145 ± 0.003	0.89 ± 0.05	0.034 ± 0.005
FT-Transformer	0.838 ± 0.005	0.669 ± 0.010	0.659 ± 0.008	0.808 ± 0.007	0.722 ± 0.012	0.139 ± 0.004	0.93 ± 0.05	0.026 ± 0.004
Proposed model	0.861 ± 0.006***	0.708 ± 0.009***	0.688 ± 0.007**	0.814 ± 0.008	0.752 ± 0.010**	0.129 ± 0.003**	0.98 ± 0.04	0.014 ± 0.003***

*Note: Values are mean ± SD across three seeds. ECE = expected calibration error. **($p<0.01$) and **($p<0.001$) versus FT-Transformer by paired participant-level bootstrap testing.

DOI: 10.54254/2755-2721/2026.37042

5.2. Ecological contributions and ablation results

Early participation carried the largest incremental information, but organizational and interpersonal features produced independent improvements after attendance behavior was retained. Removing early participation reduced AUPRC by 0.071 ± 0.008 and increased Brier score by 0.019 ± 0.004. Direct concatenation produced a smaller but significant decline, confirming that ecological grouping contributed beyond parameter count. Performance remained stable across prespecified subgroups; the largest reduction occurred in high-deprivation areas, where AUPRC was 0.681 ± 0.014 and calibration slope was 0.94 ± 0.07. Explanation rank correlation across seeds was 0.87 ± 0.03.

6. Conclusion

The study presents a structured method for identifying exercise rehabilitation discontinuation after fourteen days. By preserving individual, interpersonal, organizational, community, and early-participation features, the hierarchical model captured interactions missed by direct concatenation. Temporal and ablation tests showed improved discrimination, calibration, and specificity, with early participation contributing most strongly. Calibrated probabilities and ecological profiles support targeted follow-up based on symptoms, transport, therapist continuity, scheduling, communication, and attendance behavior. Stable subgroup performance suggests limited dependence on average service patterns. Future prospective studies should test whether ecology-matched interventions reduce discontinuation without increasing workload or inequity and assess recalibration across community centers and settings.

References

[1] Mahmood, A., Nayak, P., Deshmukh, A., English, C., Manikandan, N., Solomon, J. M., & Unnikrishnan, B. (2023). Measurement, determinants, barriers, and interventions for exercise adherence: A scoping review. *Journal of Bodywork and Movement Therapies, 33*, 95–105.

[2] Shahidi, B., Padwal, J., Lee, E., Xu, R., Northway, S., Taitano, L., Wu, T., & Raiszadeh, K. (2022). Factors impacting adherence to an exercise-based physical therapy program for individuals with low back pain. *PLOS ONE, 17*(10), e0276326.

[3] Collins, K. A., Reeves, G. R., Miller, N. H., Whellan, D. J., O'Connor, C. M., Marcus, B. H., Kitzman, D. W., Kraus, W. E., & HF-ACTION Investigators. (2023). Clinical predictors of adherence to exercise training among individuals with heart failure: The HF-ACTION study. *Journal of Cardiopulmonary Rehabilitation and Prevention, 43*(3), 205–213.

[4] Bourke, A., Niranjan, V., O'Connor, R., & Woods, C. (2022). Barriers to and motives for engagement in an exercise-based cardiac rehabilitation programme in Ireland: A qualitative study. *BMC Primary Care, 23*(1), Article 28.

[5] Bamonti, P. M., Boyle, J. T., Goodwin, C. L., Wan, E. S., Silberbogen, A. K., Finer, E. B., & Moy, M. L. (2022). Predictors of outpatient pulmonary rehabilitation uptake, adherence, completion, and treatment response among male U.S. veterans with chronic obstructive pulmonary disease. *Archives of Physical Medicine and Rehabilitation, 103*(6), 1113–1121.e1.

[6] Vancampfort, D., Van Damme, T., Brunner, E., McGrath, R. L., Hemmings, L., Guimarães, M. E., & Schuch, F. (2024). Dropout from exercise interventions in adults with fibromyalgia: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation, 105*(3), 571–579.

[7] Baidya, S., Connolly, C. J., Petersen, J. M., Baldwin, C., van den Berg, M. E. L., Harris, I., & Lewis, L. K. (2025). Barriers and facilitators to physical activity among older adults in residential aged care facilities: A mixed methods systematic review using the social ecological model. *Journal of Aging and Health, 37*(10), 702–720.

[8] Liu, D., Shin, W.-Y., Sprecher, E., Conroy, K., Santiago, O., Wachtel, G., & Santillana, M. (2022). Machine learning approaches to predicting no-shows in pediatric medical appointment. *npj Digital Medicine, 5*(1), Article 50.

DOI: 10.54254/2755-2721/2026.37042

[9] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems, 34*, 18932–18943.

[10] Huang, C., Li, S.-X., Caraballo, C., Masoudi, F. A., Rumsfeld, J. S., Spertus, J. A., Normand, S.-L. T., Mortazavi, B. J., & Krumholz, H. M. (2021). Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning. *Circulation: Cardiovascular Quality and Outcomes, 14*(10), e007526.