Supplementary Material

Table S1. Post-hoc pairwise comparisons across hospitals using Dunn's test with Holm correction for variables with significant overall Kruskal–Wallis test results.

Variable	KHUH vs KNUH	KHUH vs DUIH	KNUH vs DUIH
Age (years)	$p = 0.058$	$p < 0.001$	$p < 0.001$
Weight (kg)	$p = 0.232$	$p < 0.001$	$p = 0.066$
Total Insulin Pre-EN (units)	$p < 0.001$	$p = 0.001$	$p = 0.487$
Total Insulin Post-EN (units)	$p = 0.029$	$p = 0.411$	$p = 0.264$
Total PN Calories Pre-EN (kcal)	$p < 0.001$	$p = 0.003$	$p < 0.001$
Total PN Calories Post-EN (kcal)	$p = 0.018$	$p = 0.298$	$p = 0.292$
Total EN Calories (kcal)	$p < 0.001$	$p < 0.001$	$p < 0.001$
Avg. BGL Pre-EN (mg/dL)	$p = 0.370$	$p < 0.001$	$p < 0.001$
Avg. BGL Post-EN (mg/dL)	$p = 0.165$	$p = 0.010$	$p = 0.002$

EN: Enteral Nutrition, PN: Parenteral Nutrition, Avg. BGL: Average Blood Glucose Level, KHUH: Kyung Hee University Hospital at Gang-dong, KNUH: Kangwon National University Hospital, DUIH: Dongguk University Ilsan Hospital. Exact adjusted p-values are reported, with values smaller than 0.001 shown as $p < 0.001$.

Representative case studies

To better understand the practical behavior of the model, we analyzed six representative cases that illustrate different scenarios encountered during the PN-to-EN transition, with detailed descriptions provided in Table S2. These cases span a broad range of pre-EN BGL values (132 to 538 mg/dL), insulin sensitivity profiles (high and low IGR), and nutritional compositions, with PN contributing 3% to 76.4% of total calories. In Cases 1, 2, 4, and 5, the model's recommendations were generally aligned with the observed glycemic context, including higher recommended doses in markedly hyperglycemic cases and a lower recommended dose in a severe hypoglycemic case. Notably, Case 4 represents an extreme hyperglycemic scenario, with a pre-EN BGL of 538 mg/dL, which lies above the 99th percentile of the observed cohort distribution. To evaluate model behavior in such upper-tail cases, we examined directional consistency for events above the 99th percentile of pre-EN BGL and found that it remained high at 92.3%. Thus, although Case 4 is an extreme example and should still be interpreted with caution, the model's performance in this subgroup suggests that its recommendation is partly supported by the available subgroup analysis. In contrast, Cases 3 and 6 highlight important model limitations. In Case 3, the model recommended a high insulin dose despite a near-hypoglycemic post-EN BGL, which should be regarded as a failure case because additional insulin under these conditions could increase the risk of further glucose reduction. The model's recommendation of a high insulin dose may have been influenced by the substantial concurrent PN+EN calorie load and the high IGR signal. In Case 6, although post-EN BGL remained high, the model recommended a lower dose than the clinician-administered dose. Overall, these examples show that, while the model often generates recommendations consistent with the observed glycemic context, failure cases remain and require further refinement and prospective validation.

Implementation details

The actor was implemented as a multilayer perceptron with two hidden layers of 256 units each and ReLU activations, followed by a sigmoid output scaled to the action upper bound. The critic was implemented as a multilayer perceptron that received the concatenated state-action vector, with two hidden layers of 256 units each and ReLU activations, followed by a linear scalar Q-value output head. Both networks were trained using Adam with a learning rate of 3×10^{-4} for both the actor and the critic. Training used mini-batches of 64 for 25 epochs in each validation run. The IGR category (high vs. low) was determined using the median IGR value computed from the training fold, and the resulting threshold, which was approximately 0.3, was then applied to the corresponding held-out test data to avoid information leakage. In the final model, the reward-maximization and behavior-cloning weights in Eq. 11 were set to $\alpha = 2.5$ and $\beta = 0.9$, respectively. In Eq. 12, the action range was constrained to $[0, 30]$ as a broader upper bound above the observed dosing range.

Table S3 presents the sensitivity analysis of the trade-off coefficients α and β in Eq. 11 under the primary leave-one-subject-out evaluation protocol. Overall, directional consistency remained relatively stable across the tested parameter combinations, ranging from 77.7% to 84.3%. The best performance was achieved at $\alpha = 2.5$ and $\beta = 0.9$, which were therefore selected for the final model. In general, performance tended to decrease as β increased, suggesting that overly strong behavior-cloning regularization may constrain the policy too closely to historical clinician actions. These results indicate that the proposed

Table S2. Detailed Case Study Data.

Attribute	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Age (years)	84	58	83	76	67	45
Gender	Male	Male	Female	Female	Female	Male
Hospital	KHUH	KHUH	DUIH	DUIH	DUIH	KNUH
Height (cm)	163	162	145	148	152	165
Weight (kg)	52	34	35	51.4	80.1	59
Pre-EN BGL (mg/dL)	223	225	132	538	349	309
IGR	Low	Low	High	Low	Low	High
Total Calories (kcal)	692.52	423.61	729.75	227.38	271.80	412.34
PN Calories (kcal)	492.52	323.61	529.75	25.71	71.80	12.34
EN Calories (kcal)	200	100	200	201.67	200	400
Administered Insulin Dose (units/hr)	0.51	0.17	0.0	1.94	3.79	4.0
Post-EN BGL (mg/dL)	439	552	78	340	67	306
Recommended Insulin Dose (units/hr)	6.43	4.46	15.32	6.11	3.16	2.53

PN: Parenteral Nutrition, IGR: Insulin-to-Glucose Ratio, EN: Enteral Nutrition, Pre-EN BGL: Pre-Enteral Nutrition Blood Glucose Level and Post-EN BGL: Post-Enteral Nutrition Blood Glucose Level.

Table S3. Sensitivity analysis of the trade-off coefficients α and β in Eq. 11 under leave-one-subject-out cross-validation. Each cell reports directional consistency (%). The best performance is shown in bold.

$\alpha \backslash \beta$	0.01	0.9	5	10
0.1	81.7%	81.0%	79.1%	77.7%
0.5	81.8%	82.2%	79.3%	78.2%
2.5	81.8%	**84.3%**	79.5%	78.5%
5.0	80.8%	80.6%	80.3%	79.5%

objective is reasonably robust across a broad range of settings while still benefiting from an appropriate balance between reward maximization and behavior cloning.