

Supplementary Information

Pretraining on Adult ICU Data Accelerates Data Efficient AI Model Development in Critically Ill Children

Data access and verification

Yuxuan Liu and Jake Ormond had direct access to the data used for the analyses and verified the accuracy and completeness of the data reported.

Supplementary Results

Table S1. Feature reference table. Complete list of the 32 input features used for model training: 3 static (demographics) and 29 dynamic (vital signs, blood pressure, ventilation, blood gas, routine laboratory values). For each feature, we report unit and median (IQR) computed over all observed values in each hospital’s base harmonised dataset. Feature names follow the YAIB/ricu adult-naming convention. Four additional features available at the intersection of all four cohorts (fibrinogen, lymphocyte %, neutrophil %, platelet count) were excluded from model training due to inconsistent units or measurement scales across sites.

Feature	Description	Unit	MIMIC-IV	Unit A	Unit B	Unit C
<i>Static (demographics)</i>						
age	Age	y	65.0 (53.0–76.0)	0.7 (0.1–4.5)	2.1 (0.5–8.1)	0.7 (0.1–4.2)
sex	Sex	% male	55.8	56.9	55.9	55.3
weight	Weight	kg	77.9 (65.2–92.8)	7.1 (3.3–16.0)	12.0 (6.5–26.7)	7.2 (3.7–16.0)
<i>Vital signs</i>						
hr	Heart rate	bpm	84 (72–97)	132 (114–149)	121 (103–138)	129 (112–144)
o2sat	SpO ₂	%	97 (94–99)	98 (95–100)	97 (95–99)	97 (91–99)
resp	Respiratory rate	/min	19 (16–23)	32 (24–41)	27 (21–35)	31 (25–40)
temp	Temperature	°C	36.9 (36.6–37.2)	36.8 (36.2–37.2)	37.1 (36.6–37.6)	36.8 (36.4–37.3)
<i>Blood pressure</i>						
dbp	Diastolic BP	mmHg	61 (53–72)	60 (44–83)	48 (37–60)	46 (39–55)
map	Mean arterial pressure	mmHg	77 (68–88)	64 (53–78)	62 (51–75)	57 (49–66)
sbp	Systolic BP	mmHg	118 (104–134)	87 (72–103)	93 (79–108)	82 (71–94)
<i>Ventilation</i>						
fio2	FiO ₂	fraction	0.50 (0.40–0.50)	0.30 (0.21–0.40)	0.30 (0.25–0.40)	0.30 (0.22–0.40)
<i>Blood gas</i>						
be	Base excess	mmol/L	0.0 (–3.0–2.0)	5.6 (2.6–10.0)	4.5 (–0.5–11.4)	2.0 (–1.5–5.7)
bicar	Bicarbonate	mmol/L	24.0 (21.0–27.0)	27.1 (23.1–32.3)	29.5 (24.8–36.3)	25.5 (22.8–28.6)
cai	Ionised calcium	mmol/L	1.12 (1.06–1.17)	1.23 (1.15–1.30)	1.23 (1.16–1.29)	1.22 (1.16–1.28)
cl	Chloride	mmol/L	104 (99–108)	104 (99–109)	107 (102–111)	105 (100–110)
lact	Lactate	mmol/L	1.8 (1.2–2.9)	1.0 (0.7–1.7)	1.0 (0.7–1.4)	1.2 (0.9–1.7)
methb	Methaemoglobin	%	0.0 (0.0–1.0)	1.0 (0.7–1.3)	1.1 (0.9–1.4)	0.9 (0.7–1.2)
pco2	pCO ₂	mmHg	40.0 (35.0–47.0)	47.6 (40.2–57.5)	49.5 (40.5–61.1)	44.8 (39.8–50.7)
ph	pH	–	7.38 (7.33–7.43)	7.38 (7.32–7.43)	7.39 (7.33–7.44)	7.39 (7.34–7.43)
po2	pO ₂	mmHg	107 (80–154)	56 (41–76)	59 (44–78)	51 (37–80)
<i>Laboratory</i>						
alt	ALT	U/L	38 (18–116)	36 (23–67)	26 (15–58)	21 (14–37)
bili	Total bilirubin	mg/dL	0.9 (0.5–2.7)	1.0 (0.5–4.5)	0.4 (0.2–0.9)	0.6 (0.4–1.5)
ca	Calcium	mg/dL	8.3 (7.9–8.8)	8.8 (8.1–9.4)	8.9 (8.4–9.4)	8.9 (8.4–9.4)
crea	Creatinine	mg/dL	1.00 (0.70–1.80)	0.38 (0.26–0.64)	0.45 (0.37–0.58)	0.36 (0.25–0.59)
glu	Glucose	mg/dL	127 (105–159)	110 (83–167)	111 (88–169)	104 (88–126)
hgb	Haemoglobin	g/dL	9.7 (8.5–11.2)	9.7 (8.5–11.2)	9.7 (8.7–10.9)	11.6 (10.2–13.1)
inr_pt	INR	ratio	1.3 (1.1–1.6)	1.2 (1.1–1.4)	1.2 (1.1–1.5)	1.2 (1.1–1.4)
k	Potassium	mmol/L	4.0 (3.7–4.4)	3.9 (3.5–4.5)	4.1 (3.8–4.5)	4.1 (3.7–4.5)
mg	Magnesium	mg/dL	2.1 (1.9–2.3)	2.2 (1.9–2.4)	2.0 (1.8–2.3)	2.2 (2.0–2.5)
na	Sodium	mmol/L	139 (136–142)	140 (136–144)	141 (138–144)	139 (135–142)
phos	Phosphate	mg/dL	3.3 (2.7–4.2)	5.0 (4.0–6.0)	4.5 (3.6–5.4)	5.4 (4.5–6.4)
wbc	WBC count	10 ⁹ /L	10.5 (7.6–14.5)	10.3 (6.7–15.0)	9.4 (6.4–13.6)	11.3 (8.3–15.2)

Table S2. Clinical prediction tasks and per-cohort statistics. Cohort sizes reflect task-specific exclusion criteria and may therefore differ from Table 1 of the main text. Static classification cells report stays (event prevalence); static regression cells report stays; outcome median (IQR). Hourly-task cells report stays; prediction instances (timestep-level event prevalence or outcome median [IQR]). The extubation-readiness task was unavailable at Unit A because a derived ventilation indicator was unavailable.

Task	Definition	MIMIC-IV	Unit A	Unit B	Unit C
Mortality.	ICU death; static binary, predicted at 24 h after ICU admission.	49,526 (7.2%)	7,415 (5.9%)	2,846 (4.0%)	3,665 (3.5%)
Length of stay.	Remaining ICU stay; hourly regression.	71,965; 4,720,378; 46 (18–114)	9,400; 874,345; 89 (34–168)	3,594; 341,175; 83 (34–168)	4,771; 425,978; 85 (31–168)
AKI onset.	Creatinine-based KDIGO stage ≥ 1 onset; ¹ hourly over the first 168 h; the 12 h preceding onset are labelled positive. PICU labels use a paediatric-adapted creatinine baseline.	62,821; 2,706,946 (12.3%)	8,739; 788,735 (0.6%)	3,290; 297,050 (1.0%)	4,592; 388,048 (0.7%)
Sepsis onset.	Adult: Sepsis-3 onset; ² PICU: operational paediatric proxy adapted from Phoenix criteria; ³ hourly over the first 168 h; the 12 h preceding onset are labelled positive.	50,675; 2,433,728 (3.9%)	5,521; 311,124 (6.3%)	2,030; 99,006 (9.2%)	2,312; 136,187 (4.6%)
Kidney function.	Median serum creatinine over h24–48; static regression, predicted from h0–24 features.	33,951; 1.00 (0.70–1.80)	4,328; 0.38 (0.27–0.59)	2,174; 0.43 (0.37–0.54)	2,090; 0.37 (0.26–0.61)
Extubation readiness.	Successful extubation within next 12 h; hourly binary during first invasive ventilation episode. ⁴	15,972; 1,027,601 (14.8%)	—	1,918; 159,333 (12.2%)	1,875; 108,772 (17.9%)

Table S3. Paired zero-shot contrasts at the full paediatric cohort. The five inner-fit metrics were averaged within each outer fold before calculating the zero-shot (ZS) versus target-only (Tgt) contrast and its two-sided t_4 95% confidence interval across the five outer folds. Classification contrasts are ZS–Tgt AUROC differences. Regression contrasts are MAE reductions relative to Tgt, so positive values favour ZS; units are hours for LOS and mg/dL for kidney function. Intervals excluding zero are shown in **bold**. Fine-tuning contrasts are reported in Table 2 of the main text.

Task	Site	ZS versus Tgt [95% CI]
<i>Classification (AUROC)</i>		
Mortality	Unit A	−0.133 [−0.160, −0.107]
Mortality	Unit B	−0.223 [−0.326, −0.120]
Mortality	Unit C	−0.061 [−0.105, −0.017]
AKI	Unit A	−0.168 [−0.216, −0.120]
AKI	Unit B	−0.221 [−0.254, −0.188]
AKI	Unit C	−0.164 [−0.193, −0.135]
Sepsis	Unit A	−0.141 [−0.160, −0.122]
Sepsis	Unit B	−0.189 [−0.224, −0.154]
Sepsis	Unit C	−0.138 [−0.167, −0.109]
Extubation readiness	Unit B	−0.145 [−0.173, −0.117]
Extubation readiness	Unit C	−0.150 [−0.187, −0.114]
<i>Length of stay (MAE reduction, hours)</i>		
LOS	Unit A	−3.46 [−4.20, −2.71]
LOS	Unit B	−5.24 [−5.89, −4.59]
LOS	Unit C	−4.34 [−5.57, −3.10]
<i>Kidney function (MAE reduction, mg/dL)</i>		
Kidney fn.	Unit A	−0.090 [−0.099, −0.081]
Kidney fn.	Unit B	−0.044 [−0.048, −0.040]
Kidney fn.	Unit C	−0.121 [−0.133, −0.108]

Table S4. Site-level performance of full-cohort target-only models trained with local versus pooled PICU data. The pooled model was trained on the combined PICU development cohorts and evaluated separately at each unit; the site-specific model used only the corresponding unit's development cohort. Benefit is pooled minus site-specific AUROC for classification and site-specific minus pooled MAE for regression, so positive values favour pooled training. Values are descriptive means across the 25 nested cross-validation fits; paired uncertainty intervals were not estimated. LOS MAE is in hours and kidney-function MAE in mg/dL.

Task	Unit	Metric	Site-specific	Pooled PICU	Benefit
Mortality	A	AUROC	0.800	0.816	+0.016
Mortality	B	AUROC	0.865	0.887	+0.022
Mortality	C	AUROC	0.810	0.810	+0.000
AKI	A	AUROC	0.845	0.860	+0.015
AKI	B	AUROC	0.818	0.846	+0.028
AKI	C	AUROC	0.866	0.872	+0.006
Sepsis	A	AUROC	0.791	0.795	+0.004
Sepsis	B	AUROC	0.807	0.824	+0.017
Sepsis	C	AUROC	0.756	0.788	+0.032
LOS	A	MAE (h)	46.46	45.69	+0.77
LOS	B	MAE (h)	43.34	41.94	+1.40
LOS	C	MAE (h)	45.09	44.27	+0.82
Kidney function	A	MAE (mg/dL)	0.138	0.123	+0.015
Kidney function	B	MAE (mg/dL)	0.104	0.087	+0.017
Kidney function	C	MAE (mg/dL)	0.123	0.113	+0.010
Extubation readiness	B	AUROC	0.777	0.765	-0.012
Extubation readiness	C	AUROC	0.756	0.741	-0.015

Table S5. Target-only training and fine-tuning at every sample size. Values are mean model-specific held-out metrics across the 25 nested cross-validation fits. Tgt denotes target-only training and FT fine-tuning on the same n-stay development pool; n = 0 gives the zero-shot (ZS) result. Higher AUROC and lower MAE indicate better performance. Em-dashes indicate unavailable combinations. Length-of-stay MAE is in hours and kidney-function MAE in mg/dL.

Task and condition	Unit A		Unit B		Unit C	
	Tgt	FT	Tgt	FT	Tgt	FT
Mortality, AUROC↑						
n=0 (ZS)	0.667		0.642		0.749	
n=100	0.600	0.739	0.636	0.805	0.612	0.770
n=200	0.636	0.736	0.739	0.834	0.634	0.773
n=500	0.717	0.759	0.815	0.849	0.707	0.782
n=1,000	0.745	0.776	0.845	0.866	0.725	0.792
n=2,000	0.778	0.798	0.856	0.852	0.766	0.797
n=3,000	0.784	0.797	—	—	0.786	0.798
n=all	0.800	0.810	0.865	0.886	0.810	0.820
AKI, AUROC↑						
n=0 (ZS)	0.677		0.597		0.701	
n=100	0.598	0.733	0.609	0.640	0.733	0.778
n=200	0.654	0.757	0.661	0.669	0.752	0.785
n=500	0.713	0.791	0.717	0.733	0.801	0.809
n=1,000	0.755	0.827	0.772	0.791	0.826	0.835
n=2,000	0.778	0.835	0.818	0.825	0.844	0.858
n=3,000	0.813	0.863	—	—	0.858	0.877
n=all	0.845	0.880	0.818	0.832	0.866	0.877
Sepsis, AUROC↑						
n=0 (ZS)	0.650		0.618		0.618	
n=100	0.644	0.697	0.701	0.747	0.594	0.666
n=200	0.665	0.727	0.742	0.769	0.664	0.718
n=500	0.732	0.746	0.784	0.796	0.703	0.736
n=1,000	0.746	0.769	0.802	0.812	0.738	0.757
n=2,000	0.775	0.783	—	—	—	—
n=3,000	0.782	0.788	—	—	—	—
n=all	0.791	0.795	0.807	0.809	0.756	0.776
LOS, MAE (h)↓						
n=0 (ZS)	49.92		48.58		49.43	
n=100	58.07	48.86	54.12	45.01	57.63	47.12
n=200	56.25	47.92	52.05	43.02	56.08	46.25
n=500	55.58	47.16	51.60	41.55	54.84	45.33
n=1,000	52.93	46.69	47.67	40.76	50.91	44.48
n=2,000	49.55	46.25	44.53	39.30	46.94	43.83
n=3,000	48.35	45.78	42.93	38.05	45.67	43.08
n=all	46.46	44.67	43.34	38.64	45.09	42.57
Kidney fn., MAE (mg/dL)↓						
n=0 (ZS)	0.228		0.148		0.244	
n=100	0.286	0.166	0.162	0.110	0.187	0.130
n=200	0.280	0.153	0.148	0.098	0.177	0.141
n=500	0.253	0.139	0.145	0.090	0.160	0.142
n=1,000	0.204	0.126	0.128	0.086	0.135	0.125
n=all	0.138	0.120	0.104	0.081	0.123	0.102
Extub. ready, AUROC↑						
n=0 (ZS)	—		0.632		0.606	
n=100	—	—	0.671	0.725	0.636	0.665
n=200	—	—	0.713	0.745	0.672	0.695
n=500	—	—	0.746	0.766	0.707	0.731
n=1,000	—	—	0.763	0.781	0.730	0.751
n=all	—	—	0.777	0.791	0.756	0.779

Table S6. Pooled-PICU target-only training and fine-tuning at every sample size. Values are mean model-specific held-out metrics across the 25 nested cross-validation fits, calculated as micro metrics over the combined held-out observations from the three PICUs. Tgt denotes target-only training and FT fine-tuning; n is the total number of pooled paediatric development stays. Higher AUROC and lower MAE indicate better performance. Em-dashes indicate sample sizes not evaluated for that task. LOS MAE is in hours and kidney-function MAE in mg/dL.

Task	Method	n = 100	n = 200	n = 500	n = 1,000	n = 2,000	n = 3,000	n = 5,000	n = 10,000	n = all
Mortality, AUROC	Tgt	0.621	0.679	0.729	0.755	0.788	0.808	0.817	0.830	0.833
	FT	0.761	0.773	0.781	0.789	0.806	0.821	0.825	0.836	0.837
AKI, AUROC	Tgt	0.585	0.621	0.736	0.766	0.788	0.805	0.829	0.845	0.861
	FT	0.677	0.716	0.777	0.810	0.836	0.846	0.859	0.865	0.879
Sepsis, AUROC	Tgt	0.631	0.694	0.734	0.756	0.782	0.791	0.801	—	0.806
	FT	0.711	0.741	0.763	0.774	0.791	0.797	0.804	—	0.808
LOS, MAE (h)	Tgt	58.47	57.08	54.66	51.97	48.63	47.38	46.25	45.10	44.51
	FT	49.85	48.20	47.19	46.51	45.21	44.97	44.37	43.38	43.01
Kidney function, MAE (mg/dL)	Tgt	0.243	0.231	0.197	0.184	0.133	0.124	0.116	—	0.111
	FT	0.141	0.139	0.120	0.115	0.109	0.107	0.106	—	0.103
Extubation readiness, AUROC	Tgt	0.614	0.640	0.695	0.711	0.731	0.748	—	—	0.758
	FT	0.657	0.695	0.720	0.741	0.758	0.773	—	—	0.784

Table S7. Precision–recall performance at the full paediatric cohort. AUPRC for zero-shot (ZS), fine-tuning (FT), and target-only training (Tgt), reported as mean model-specific held-out metrics across the 25 nested cross-validation fits. Event rate is the corresponding mean held-out positive prevalence and the no-skill AUPRC reference.

Task	Site	Event rate (%)	ZS	FT	Tgt
Mortality	Unit A	5.9	0.141	0.289	0.276
Mortality	Unit B	3.9	0.155	0.378	0.353
Mortality	Unit C	3.3	0.143	0.204	0.174
AKI	Unit A	0.6	0.029	0.064	0.061
AKI	Unit B	1.0	0.019	0.058	0.056
AKI	Unit C	0.7	0.041	0.082	0.090
Sepsis	Unit A	6.1	0.110	0.224	0.221
Sepsis	Unit B	9.0	0.130	0.280	0.280
Sepsis	Unit C	4.7	0.083	0.154	0.149
Extub. ready	Unit B	12.4	0.182	0.340	0.326
Extub. ready	Unit C	18.5	0.266	0.460	0.405

Table S8. Predicted risk and Brier score at the full paediatric cohort. Fine-tuned (FT) and target-only (Tgt) models underwent intercept-only recalibration fitted on validation predictions and applied unchanged to held-out test predictions. Zero-shot (ZS) predictions were not recalibrated. Values are means across the 25 nested cross-validation fits.

Task	Site	Event rate (%)	Mean predicted risk (%)			Brier score		
			ZS	FT	Tgt	ZS	FT	Tgt
Mortality	Unit A	5.89	65.42	5.85	5.84	0.4198	0.0516	0.0516
Mortality	Unit B	3.91	61.26	3.99	3.96	0.3849	0.0329	0.0336
Mortality	Unit C	3.34	61.37	3.41	3.41	0.3934	0.0309	0.0311
AKI	Unit A	0.57	41.70	0.58	0.58	0.1825	0.0056	0.0056
AKI	Unit B	0.96	37.30	0.96	0.97	0.1478	0.0093	0.0093
AKI	Unit C	0.72	40.19	0.77	0.76	0.1733	0.0070	0.0070
Sepsis	Unit A	6.07	48.40	6.02	6.01	0.2459	0.0539	0.0540
Sepsis	Unit B	8.99	48.05	8.84	8.80	0.2412	0.0755	0.0754
Sepsis	Unit C	4.72	46.27	5.05	5.03	0.2259	0.0433	0.0434
Extub. ready	Unit B	12.44	32.51	12.40	12.34	0.1476	0.0988	0.1007
Extub. ready	Unit C	18.48	33.49	18.23	18.30	0.1701	0.1338	0.1385

Table S9. Diagnostic comparison of adaptation strategies at the full paediatric cohort. Classification columns compare source-normalised zero-shot, external-PICU-normalised zero-shot, linear probing and full fine-tuning. Regression columns compare zero-shot, target-only training, linear probing and full fine-tuning. Values are mean model-specific held-out AUROC or MAE across the 25 nested cross-validation fits. External-PICU normalisation uses statistics from the other two PICUs without evaluated-site data; linear probing freezes the adult encoder and trains a new output layer on labelled target data.

Task	Site	Source-normalised zero-shot	External-PICU-normalised zero-shot	Linear probing	Full fine-tuning
<i>Classification — mean AUROC</i>					
Mortality	Unit A	0.667	0.687	0.771	0.810
Mortality	Unit B	0.642	0.746	0.801	0.886
Mortality	Unit C	0.749	0.787	0.798	0.820
AKI	Unit A	0.677	0.741	0.835	0.880
AKI	Unit B	0.597	0.601	0.777	0.832
AKI	Unit C	0.701	0.699	0.819	0.877
Sepsis	Unit A	0.650	0.664	0.775	0.795
Sepsis	Unit B	0.618	0.607	0.794	0.809
Sepsis	Unit C	0.618	0.631	0.748	0.776
Extubation readiness	Unit B	0.632	0.637	0.763	0.791
Extubation readiness	Unit C	0.606	0.602	0.752	0.779
<i>Regression — mean MAE (LOS in hours; kidney creatinine in mg/dL)</i>					
		Zero-shot	Target-only	Linear probing	Full fine-tuning
Length of stay	Unit A	49.9	46.5	48.5	44.7
Length of stay	Unit B	48.6	43.3	45.5	38.6
Length of stay	Unit C	49.4	45.1	47.1	42.6
Kidney function	Unit A	0.228	0.138	0.127	0.120
Kidney function	Unit B	0.148	0.104	0.088	0.081
Kidney function	Unit C	0.244	0.123	0.111	0.102

Supplementary References

1. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. *Kidney Int. Suppl.* **2**, 1–138 (2012).
2. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* **315**, 801–810 (2016).
3. Schlapbach, L. J. et al. International consensus criteria for pediatric sepsis and septic shock. *JAMA* **331**, 665–674 (2024).
4. Liu, Y. et al. Dual-phase machine learning framework for paediatric ventilator liberation: Forecasting extubation readiness and nowcasting extubation outcomes. *BMJ Digit. Health AI* **1**, e000138 (2025).