Supplementary Material for:

Deep learning models for ICU readmission prediction: A systematic review and meta-analysis

Emanuele Koumantakis[1], Konstantina Remoundou[2,3], Nicoletta Colombi[4], Carmen Fava[5], Ioanna Roussaki[2,3], Alessia Visconti[1, ‡,*], Paola Berchialla[1, ‡]

[1]Center for Biostatistics, Epidemiology, and Public Health, Department of Clinical and Biological Sciences, University of Turin, Turin, Italy

[2]School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

[3]Institute of Communication and Computer Systems, Athens, Greece

[4]Federated Library of Medicine "F. Rossi", University of Turin, Turin, Italy

[5]Department of Clinical and Biological Sciences, University of Turin, Turin, Italy

*Corresponding author

‡These authors jointly supervised this work.

Supplementary Tables

Supplementary Table 1. Search strategies for each database. The table includes the search strings, the fields searched, and the number of results retrieved on March 4th, 2025. The gray lines highlight the query and the number of documents that were selected for screening.

Database	Step	Queries	Results
PubMed	#1	"Intensive Care Units"[Mesh]	115,009
	#2	"intensive care"[tiab] OR ICU [tiab] OR ICUs[tiab] OR "coronary care unit*"[tiab] OR "respiratory care unit*"[tiab]	26,0187
	#3	#1 OR #2	28,9182
	#4	"Neural Networks, Computer"[Mesh]	78,727
	#5	"deep learning"[tiab] OR "deep machine learning"[tiab] OR "deep ML"[tiab] OR "deep reinforcement learning"[tiab] OR "deep RL"[tiab] OR "hierarchical learning"[tiab] OR "neural network*"[tiab] OR autoencoder*[tiab] OR auto-encoder*[tiab] OR "long short-term memory"[tiab] OR "conditional random field*"[tiab] OR transformer*[tiab] OR "deep belief network"[tiab]	186,804
	#6	"Machine Learning"[Mesh:noexp] OR "Machine Learning Algorithms"[Mesh:noexp]	50,009
	#7	"machine learning"[tiab]	140,393

	#8	#4 OR #5 OR #6 OR #7	305,400
	#9	"Patient Readmission"[Mesh]	25,102
	#10	readmi*[tiab] OR re-admi*[tiab] OR rehospital*[tiab] OR re-hospital*[tiab]	69,068
	#11	#9 OR #10	73,674
	#12	#3 AND #8 AND #11	88
Embase	#1	intensive care unit'/exp OR 'intensive care'/exp	1200626
	#2	'intensive care':ti,ab,kw OR ICU:ti,ab,kw OR ICUs:ti,ab,kw OR 'coronary unit*':ti,ab,kw OR 'respiratory care unit*':ti,ab,kw	415,810
	#3	#1 OR #2	1,292,732
	#4	'artificial neural network'/exp OR 'deep learning'/exp OR 'deep learning algorithm'/exp OR 'deep learning model'/exp OR 'conditional random field'/exp OR 'transformer'/exp OR 'transformer model'/exp	169,353
	#5	'deep learning':ti,ab,kw OR 'deep machine learning':ti,ab,kw OR 'deep ML':ti,ab,kw OR 'deep reinforcement learning':ti,ab,kw OR 'deep RL':ti,ab,kw OR 'hierarchical learning':ti,ab,kw OR 'neural network*':ti,ab,kw OR autoencoder*:ti,ab,kw OR auto-encoder*:ti,ab,kw OR 'long short-term memory':ti,ab,kw OR 'conditional random field*':ti,ab,kw OR transformer*:ti,ab,kw OR 'deep belief network*':ti,ab,kw	212,984
	#6	'machine learning'/de OR 'machine learning algorithm'/de	150,275

	#7	'machine learning':ti,ab,kw	159,361
	#8	#4 OR #5 OR #6 OR #7	383,673
	#9	'hospital readmission'/exp	116,004
	#10	readmi*:ti,ab,kw OR re-admi*:ti,ab,kw OR rehospital*:ti,ab,kw OR re-hospital*:ti,ab,kw	125,997
	#11	#9 OR #10	154,400
	#12	#3 AND #8 AND #11	279
Scopus	#1	TITLE-ABS-KEY("intensive care" OR ICU OR ICUs OR "coronary care unit*" OR "respiratory care unit*")	485,155
	#2	TITLE-ABS-KEY("deep learning" OR "deep machine learning" OR "deep ML" OR "deep reinforcement learning" OR "deep RL" OR "hierarchical learning" OR "neural network*" OR autoencoder* OR auto-encoder* OR "long short-term memory" OR "conditional random field*" OR transformer* OR "deep belief network" OR "machine learning")	2,133,110
	#3	TITLE-ABS-KEY(readmi* OR re-admi* OR rehospital* OR re-hospital*)	104,951
	#4	#1 AND #2 AND #3	210
Web of Science	#1	TS=("intensive care" OR ICU OR ICUs OR "coronary care unit*" OR "respiratory care unit*")	282,601

	#2	TS=("deep learning" OR "deep machine learning" OR "deep ML" OR "deep reinforcement learning" OR "deep RL" OR "hierarchical learning" OR "neural network*" OR autoencoder* OR auto-encoder* OR "long short-term memory" OR "conditional random field*" OR transformer* OR "deep belief network" OR "machine learning")	1,421,395
	#3	TS=(readmi* OR re-admi* OR rehospital* OR re-hospital*)	76,360
	#4	#1 AND #2 AND #3	117

Supplementary Table 2. Summary of the included studies. The table summarizes, for each study, whether an ICU subpopulation was used, and, if so, which one, the types of the predictors used (*i.e.,* static, dynamic, unstructured text), the definition and timeframe of the ICU readmission outcome, and the source of the data, including the datasets used for training and for internal and/or external validation, as well as the internal validation approach. For each dataset, the table also reports the country in which the data were collected and the sample size (N).

Author (year)	**ICU subpopulation**	**Type(s) of predictors**	**ICU readmission outcome and timeframe**	**Data source**		
				Training (country; N)	**Internal validation approach**	**External Validation (country; N)**
Agmon et al. (2024) [1]	No	Static; dynamic	30-day readmission and mortality	MIMIC-III (USA; 38,552)	Training (80%) and testing (20%) sets	-
Barbieri et al. (2020) [2]	No	Static; dynamic	30-day readmission	MIMIC-III (USA; 45,298)	Training (90%) and testing (10%) sets	-
Cao et al. (2023) [3]	No	Static; dynamic	Readmission within hospital stay	MIMIC-III (USA; 14911) and eICU (USA; 14173)	Not clear	-
Chiu et al. (2024) [4]	No	Static; dynamic; unstructured	30-day readmission	MIMIC-III (USA; 36,232)	Training (80%) and testing (20%) sets	-

Author (year)	ICU subpopulation	Type(s) of predictors	ICU readmission outcome and timeframe	Data source		
				Training (country; N)	Internal validation approach	External Validation (country; N)
		text				
Darabi et al. (2020) [5]	No	Static; dynamic; unstructured text	30-day readmission	MIMIC-III (USA; 38,402)	Seven-fold cross-validation	-
Deng et al. (2022) [6]	No	Static; dynamic	30-day readmission	MIMIC-IV (USA; 36,180)	Training (80%) and testing (20%) sets	-
El Hassouny et al. (2024) [7]	No	Static; dynamic	30-day readmission	Multicentric US ICU dataset (USA; 69,540)	Training (50%) and testing (50%) sets	-
Ibrahim et al. (2022) [8]	Pneumonia; CKD	Static; dynamic	Readmission within different timeframes: 5 days; 7 days; 14 days; 30 days	MIMIC-III (USA; pneumonia: 2,798, CKD: 2,822)	Three-fold cross-validation	-
Kessler et al. (2023) [9]	CSRU or CCU; all ICUs	Static; dynamic	48-hour readmission and mortality	MIMIC-III (USA; 40,663) CSRU/CCU: 12,797	Training (80%) and testing (20%) sets	-

Author (year)	ICU subpopulation	Type(s) of predictors	ICU readmission outcome and timeframe	Data source		
				Training (country; N)	Internal validation approach	External Validation (country; N)
Khodadadi et al. (2023) [10]	No	Static	Readmission within hospital stay	eICU (USA; 41,026)	Training (75%) and testing (25%) sets	-
Lim et al. (2025) [11]	No	Static; dynamic	Readmission and mortality within hospital stay (all lengths of stay after ICU discharge; within 48 hours; after 48 hours)	SNUH (South Korea; 70,842)	Chronological split: training: data collected between 2007 and 2018; testing: data collected between 2019 and 2021	MIMIC-III (USA; 43,237) and eICU (USA; 90,271)
Lin et al. (2019) [12]	No	Static; dynamic	30-day readmission and mortality	MIMIC-III (USA; 48;393)	Training (80%), validation (10%), and testing (10%) sets	-
Lu et al. (2020) [13]	No	Static; dynamic	30-day readmission	MIMIC-III (USA; 48,410)	Training (85%), validation (5%), and testing (10%) sets	-
Lu et al. (2021) [14]	No	unstructured	30-day readmission	MIMIC-III (USA; no	Training (80%),	-

Author (year)	ICU subpopulation	Type(s) of predictors	ICU readmission outcome and timeframe	Data source		
				Training (country; N)	Internal validation approach	External Validation (country; N)
		text		information)	validation (10%), and testing (10%) sets	
Moazemi et al. (2022) [15]	CCU	Static; dynamic	30-day readmission and mortality	MIMIC-III (USA; 11,513)	training and testing sets (no information)	Universitätsklinikum Düsseldorf ICU (Germany; 502)
Niu et al. (2023) [16]	No	Static; dynamic	Readmission within different timeframes: 5 days; 10 days; 15 days; 30 days	MIMIC-III (USA; 133,383)	No information	-
Pei et al. (2021) [17]	No	Static; dynamic	30-day readmission	MIMIC-III (USA; 45,298)	Training and validation (90%) and testing (10%) sets	-
Pishgar et al. (2022) [18]	HF	Static; dynamic	30-day readmission	MIMIC-III (USA; 3,411)	Training (73.1%), validation (12.9%), and testing (14%) sets	-
Rasheed et al. (2024) [19]	No	Static; dynamic	30-day readmission	MIMIC-III (USA;	Training (70%),	-

Author (year)	ICU subpopulation	Type(s) of predictors	ICU readmission outcome and timeframe	Data source		
				Training (country; N)	Internal validation approach	External Validation (country; N)
			and mortality	38,348)	validation (15%), and testing (15%) sets	
Sheetrit et al. (2023) [20]	No	Static; dynamic; unstructured text	30-day readmission	MIMIC-III (USA; 14,837)	Five-fold cross-validation	-
Shickel et al. (2022) [21]	No	Static; dynamic	Readmission and mortality within hospital stay	University of Florida Integrated Data Repository (USA; 73,190)	Chronological split: training earliest 80%, and testing most recent 20%	-
Sun et al. (2024) [22]	No	Dynamic; unstructured text	Readmission within hospital stay	eICU (USA; no information)	Train (70%), validation (15%) and test sets (15%)	-
Wang et al. (2023) [23]	No	Static; dynamic	30-day readmission and mortality	MIMIC-III (USA; 38,023)	Train (60%), validation (10%) and test sets (30%)	-
Zebin and Chaussalet (2019)	No	Static; dynamic	30-day readmission	MIMIC-III (USA;	Train (80%), validation	-

Author (year)	ICU subpopulation	Type(s) of predictors	ICU readmission outcome and timeframe	Data source		
				Training (country; N)	Internal validation approach	External Validation (country; N)
[24]			and mortality	48,393)	(10%) and test sets (10%)	

Abbreviations: CKD, Chronic Kidney Disease; CSRU, Cardiosurgery recovery unit; CCU, Coronary care unit; HF, Heart failure.

Supplementary Table 3. Summary of the described models. The table summarizes, for each study, the (sub)population, the timeframe of the ICU readmission outcome, and the Deep Learning (DL) architecture(s) used. For each model (defined as a combination of subpopulation, timeframe, and architecture), the table reports the performance obtained during internal and external validation (when performed, as indicated by the notes) along with the corresponding 95% confidence interval (CI) or standard error (SE), when available.

Author (year)	**Population**	**Readmission timeframe**	**DL architecture**	**Performance metric**	**Value**	**95% CI/SE**
Agmon et al. (2024) [1]	All ICUs/diagnoses	30 days	TeDi-BERT + LSTM	AUROC	0.65	-
Barbieri et al. (2020) [2]	All ICUs/diagnoses	30 days	ODE + RNN + Attention	AUROC	0.739	0.736 - 0.741
				Precision	0.314	0.306 - 0.321
				F1 Score	0.376	0.371 - 0.381
				Sensitivity	0.685	0.666 - 0.704
				Specificity	0.677	0.658 - 0.696
	All ICUs/diagnoses	30 days	ODE + RNN	AUROC	0.739	0.737 - 0.742

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Precision	0.331	0.323 - 0.339
				F1 Score	0.372	0.367 - 0.377
				Sensitivity	0.672	0.659 - 0.686
				Specificity	0.697	0.683 - 0.711
	All ICUs/diagnoses	30 days	RNN (ODE time decay) + Attention	AUROC	0.743	0.741 - 0.746
				Precision	0.316	0.307 - 0.324
				F1 Score	0.375	0.370 - 0.379
				Sensitivity	0.648	0.641 - 0.656
				Specificity	0.733	0.726 - 0.739
	All ICUs/diagnoses	30 days	RNN (ODE time decay)	AUROC	0.741	0.738 - 0.744
				Precision	0.3	0.293 - 0.308
				F1 Score	0.372	0.367 - 0.376

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Sensitivity	0.71	0.698 - 0.722
				Specificity	0.667	0.655 - 0.679
	All ICUs/diagnoses	30 days	RNN (exp time decay) + Attention	AUROC	0.748	0.745 - 0.751
				Precision	0.32	0.312 - 0.328
				F1 Score	0.377	0.372 - 0.382
				Sensitivity	0.704	0.692 - 0.715
				Specificity	0.68	0.668 - 0.692
	All ICUs/diagnoses	30 days	RNN (exp time decay)	AUROC	0.735	0.732 - 0.738
				Precision	0.304	0.297 - 0.311
				F1 Score	0.368	0.363 - 0.373
				Sensitivity	0.707	0.700 - 0.714
				Specificity	0.67	0.663 - 0.676

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	30 days	RNN (concatenated Δtime) + Attention	AUROC	0.741	0.739 - 0.744
				Precision	0.312	0.303 - 0.320
				F1 Score	0.368	0.363 - 0.372
				Sensitivity	0.687	0.680 - 0.695
				Specificity	0.688	0.681 - 0.696
	All ICUs/diagnoses	30 days	RNN (concatenated Δtime)	AUROC	0.739	0.737 - 0.742
				Precision	0.311	0.303 - 0.320
				F1 Score	0.364	0.359 - 0.369
				Sensitivity	0.698	0.692 - 0.704
				Specificity	0.688	0.684 - 0.693
	All ICUs/diagnoses	30 days	ODE + Attention	AUROC	0.717	0.714 - 0.720
				Precision	0.294	0.285 - 0.302

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				F1 Score	0.333	0.328 - 0.339
				Sensitivity	0.776	0.768 - 0.784
				Specificity	0.554	0.548 - 0.560
	All ICUs/diagnoses	30 days	Attention (concatenated time)	AUROC	0.711	0.709 - 0.714
				Precision	0.286	0.277 - 0.295
				F1 Score	0.33	0.325 - 0.334
				Sensitivity	0.7	0.686 - 0.714
				Specificity	0.614	0.601 - 0.628
	All ICUs/diagnoses	30 days	MCE + RNN + Attention	AUROC	0.736	0.734 - 0.739
				Precision	0.317	0.308 - 0.325
				F1 Score	0.373	0.369 - 0.378
				Sensitivity	0.63	0.622 - 0.638

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Specificity	0.744	0.738 - 0.749
	All ICUs/diagnoses	30 days	MCE + RNN + Attention	AUROC	0.727	0.724 - 0.730
				Precision	0.298	0.291 - 0.306
				F1 Score	0.361	0.357 - 0.366
				Sensitivity	0.654	0.645 - 0.663
				Specificity	0.706	0.697 - 0.715
	All ICUs/diagnoses	30 days	MCE + Attention	AUROC	0.689	0.686 - 0.692
				Precision	0.269	0.261 - 0.278
				F1 Score	0.312	0.308 - 0.316
				Sensitivity	0.686	0.676 - 0.695
				Specificity	0.616	0.607 - 0.625
Cao et al. (2023) [3]		Within hospital stay	MMMGCL	Accuracy	0.7772	0.005

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
	Medical ICU(1)			PRAUC	0.4131	0.026
	CCU(1)	Within hospital stay	MMMGCL	Accuracy	0.8875	0.015
				PRAUC	0.1608	0.016
	CSRU(1)	Within hospital stay	MMMGCL	Accuracy	0.9461	0.004
				PRAUC	0.0908	0.010
	Medical ICU(2)	Within hospital stay	MMMGCL	Accuracy	0.8887	0.001
				PRAUC	0.2117	0.011
Chiu et al. (2024) [4]	All ICUs/diagnoses	30 days	LSTM	AUROC	0.7532	0.7490 - 0.7574
				Precision	0.8802	0.0082
				F1 Score	0.8558	0.0065
				Accuracy	0.8857	0.0027
				Recall	0.7834	0.0069

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	30 days	BERTopic-LSTM	AUROC	0.7994	0.7957 - 0.8031
				Precision	0.9355	0.0072
				F1 Score	0.8788	0.0063
				Accuracy	0.9316	0.0024
				Recall	0.8286	0.0088
Darabi et al. (2020) [5]	All ICUs/diagnoses	30 days	BioBERT + GRU	AUROC	0.6742	0.021
				PRAUC	0.6803	0.015
Deng et al. (2022) [6]	All ICUs/diagnoses	30 days	RNN	AUROC	0.625	0.008
				Sensitivity	0.658	0.036
				Specificity	0.567	0.039
				Brier Score	0.105	0.001
	All ICUs/diagnoses	30 days	GRU	AUROC	0.631	0.011

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Sensitivity	0.652	0.083
				Specificity	0.541	0.072
				Brier Score	0.105	0.002
	All ICUs/diagnoses	30 days	LSTM	AUROC	0.635	0.018
				Sensitivity	0.691	0.064
				Specificity	0.524	0.061
				Brier Score	0.104	0.009
El Hassouny et al. (2024) [7]	All ICUs/diagnoses	30 days	BOPS + OIMP + GSP-based algorithm (EPRIC)	Not provided	-	-
Ibrahim et al. (2022) [8]	Pneumonia	5, 7, 14, and 30 days	LSTM + gradient boost trees	AUROC(3)	0.87	0.825 - 0.926
				Precision(3)	0.873	0.737 - 0.931
				F1 Score(3)	0.881	0.774 - 0.951

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				PRAUC(3)	0.881	0.727 - 0.967
				Recall(3)	0.869	0.825 - 0.926
	Chronic Kidney Disease	5, 7, 14, and 30 days	LSTM + gradient boost trees	AUROC(3)	0.895	0.826 - 0.945
				Precision(3)	0.868	0.760 - 0.905
				F1 Score(3)	0.919	0.781 - 0.985
				PRAUC(3)	0.895	0.756 - 0.998
				Recall(3)	0.889	0.802 - 0.985
Kessler et al. (2023) [9]	CSRU/CCU	2 days	LSTM	AUROC	0.777	-
				Precision	0.268	-
				F1 Score	0.372	-
				PRAUC	0.259	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Recall	0.607	-
	All ICUs/diagnoses	2 days	LSTM	AUROC	0.861	-
				PRAUC	0.706	-
Khodadadi et al. (2023) [10]	All ICUs/diagnoses	Within hospital stay	Deep Forests	AUROC	0.869	-
				PRAUC	0.5952	-
Lim et al. (2025) [11]	All ICUs/diagnoses	Within hospital stay	LSTM + LightGBM	AUROC	0.82	0.808 - 0.832
				AUROC(4)	0.768	0.748 - 0.787
				AUROC(5)	0.725	0.712 - 0.739
	All ICUs/diagnoses	2 days (within hospital stay)	LSTM + LightGBM	AUROC	0.771	0.743 - 0.798
				AUROC(4)	0.726	0.687 - 0.764
				AUROC(5)	0.686	0.665 - 0.707
	All ICUs/diagnoses		LSTM + LightGBM	AUROC	0.834	0.821 - 0.846

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
		More than 2 days (within hospital stay)		AUROC(4)	0.782	0.760 - 0.804
				AUROC(5)	0.759	0.742 - 0.776
Lin et al. (2019) [12]	All ICUs/diagnoses	30 days	CNN	AUROC	0.784	0.773 - 0.794
				Recall	0.735	0.676 - 0.794
	All ICUs/diagnoses	30 days	CNN + LSTM	AUROC	0.787	0.775 - 0.799
				Recall	0.71	0.648 - 0.771
	All ICUs/diagnoses	30 days	LSTM + CNN	AUROC	0.791	0.782 - 0.800
				Recall	0.742	0.718 - 0.766
Lu et al. (2020) [13]	All ICUs/diagnoses	30 days	GCN + MRGAE	AUROC	0.7807	-
				Precision	0.3687	-
				Accuracy	0.7094	-
				PRAUC	0.477	-

Author (year)	**Population**	**Readmission timeframe**	**DL architecture**	**Performance metric**	**Value**	**95% CI/SE**
				Recall	0.7259	-
Lu et al. (2021) [14]	All ICUs/diagnoses	30 days	GCN + LSTM	AUROC	0.825	-
				PRAUC	0.632	-
				Recall (at 80% precision)	0.319	-
Moazemi et al. (2022) [15]	Cardiovascular care units	30 days	LSTM (cropped)	AUROC	0.877	-
				Precision	0.389	-
				Accuracy	0.799	-
				PRAUC	0.666	-
				Recall	0.694	-
				AUROC(4)	0.796	-
				Precision(4)	0.35	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Accuracy(4)	0.74	-
				PRAUC(4)	0.538	-
				Recall(4)	0.575	-
				AUROC(6)	0.554	-
				Precision(6)	0.318	-
				Accuracy(6)	0.514	-
				PRAUC(6)	0.223	-
				Recall(6)	0.06	-
	Cardiovascular care units	30 days	LSTM (uncropped)	AUROC	0.859	-
				Precision	0.389	-
				Accuracy	0.778	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				PRAUC	0.636	-
				Recall	0.649	-
				AUROC(4)	0.828	-
				Precision(4)	0.477	-
				Accuracy(4)	0.766	-
				PRAUC(4)	0.571	-
				Recall(4)	0.628	-
				AUROC(6)	0.517	-
				Precision(6)	0.196	-
				Accuracy(6)	0.498	-
				PRAUC(6)	0.221	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Recall(6)	0.22	-
Niu et al. (2023) [16]	All ICUs/diagnoses	5 days	MP-ROM + RNN	AUROC	0.737	-
				Precision	0.405	-
				F1 Score	0.463	-
	All ICUs/diagnoses	10 days	MP-ROM + RNN	AUROC	0.743	-
				Precision	0.538	-
				F1 Score	0.557	-
	All ICUs/diagnoses	15 days	MP-ROM + RNN	AUROC	0.736	-
				Precision	0.598	-
				F1 Score	0.586	-
	All ICUs/diagnoses	30 days	MP-ROM + RNN	AUROC	0.726	-
				Precision	0.666	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				F1 Score	0.635	-
				Sensitivity	0.588	-
				Specificity	0.77	-
Pei et al. (2021) [17]	All ICUs/diagnoses	30 days	Graph Attention + RNN-based ODE	AUROC	0.786	- (7)
				Precision	0.375	0.366 - 0.384
				F1 Score	0.422	0.416 - 0.427
				Sensitivity	0.74	0.734 - 0.746
				Specificity	0.707	0.7 - 0.713
Pishgar et al. (2022) [18]	Heart failure	30 days	Decay Replay Mining + NN	AUROC	0.93	0.898 - 0.960
				Precision	0.886	-
				F1 Score	0.8	-
				Sensitivity	0.805	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
				Accuracy	0.841	-
Rasheed et al. (2024) [19]	All ICUs/diagnoses	30 days	CNN	AUROC	0.882	-
				Precision	0.788	-
				Accuracy	0.816	-
				PRAUC	0.866	-
				Recall	0.863	-
Sheetrit et al. (2023) [20]	All ICUs/diagnoses	30 days	GRN-based bidirectional RNN	AUROC	0.6235	0.0144
				Precision	0.1635	0.016
				F1 Score	0.2517	0.0126
				PRAUC	0.1598	0.0105
				Recall	0.5604	0.0525
Shickel et al. (2022) [21]	All ICUs/diagnoses	Within hospital stay	Transformer	AUROC	0.843	-

Author (year)	Population	Readmission timeframe	DL architecture	Performance metric	Value	95% CI/SE
Sun et al. (2024) [22]	All ICUs/diagnoses	Within hospital stay	Code-Text cross-modal Contrastive Learning (CTCL)	AUROC	0.8531	0.0037
				F1 Score	0.7933	0.0037
				PRAUC	0.8851	0.0042
Wang et al. (2023) [23]	All ICUs/diagnoses	30 days	Generated causal graph paths + BERT	AUROC	0.7933	0.7725 - 0.8141
				Precision	0.8233	0.8141 - 0.8325
				F1 Score	0.6356	0.6284 - 0.6428
				Accuracy	0.8516	0.8441 - 0.8591
				PRAUC	0.5388	0.5156 - 0.5620
Zebin and Chaussalet (2019) [24]	All ICUs/diagnoses	30 days	LSTM + CNN	AUROC	0.821	-
				Precision	0.922	-
				Accuracy	0.731	-
				Recall	0.742	-

Notes

(1) Internal validation on MIMIC-III

(2) Internal validation on eICU

(3) The performance of the model was averaged over the different readmission timeframes

(4) External validation on MIMIC-III

(5) External validation on eICU

(6) External validation on the Universitätsklinikum Düsseldorf ICU dataset

(6) Not correctly reported

Abbreviations: AUROC, area under the receiver operating characteristic; BOPS, budgeted online patient selection; CCU, coronary care unit; CI, confidence interval; CNN, convolutional neural network; CSRU, cardiosurgery recovery unit; GRU, gated recurrent unit; GSP, generalized secretary problem; LSTM, long short-term memory; MCE, medical concept embeddings; MMMGCL, multi-view graph contrastive learning; MP-ROM, multi-task learning with Pearson and RNN-based neural ordinary ODEs model; NN, neural network; ODE, ordinary differential equation; OIMP, optimal interval of misclassified patients; PRAUC, area under the precision-recall curve; RNN, recurrent neural network.

Supplementary Table 4. PROBAST assessment of the included studies. The table summarizes, for each study, the score assigned to each domain of the Prediction model Risk Of Bias ASsessment Tool (PROBAST), along with the overall judgement of risk of bias and applicability. According to PROBAST guidance [doi: 10.7326/M18-1376], the overall risk of bias was rated high if at least one domain was rated high, and rated low only if all domains were rated low. If one or more domains were rated unclear and none was rated high, the overall judgement was rated unclear.

Author (year)	**Risk of Bias**					**Applicability**			
	Participants	**Predictors**	**Outcome**	**Analysis**	**Overall**	**Participants**	**Predictors**	**Outcome**	**Overall**
Agmon et al. (2024) [1]	Low	Low	High(1)	Unclear(2)	High	Low	Low	High(1)	High
Barbieri et al. (2020) [2]	Low	Low	Low	Low	Low	Low	Low	Low	Low
Cao et al. (2023) [3]	Low	Low	Low	High(3)	High	Low	Low	Low	Low
Chiu et al. (2024) [4]	Low	Low	Low	High(9)	High	Low	Low	Low	Low
Darabi et al. (2020) [5]	Low	Low	Unclear(4)	Low	Unclear	Low	Low	Low	Low
Deng et al. (2022) [6]	Low	Low	Low	Low	Low	Low	Low	Low	Low
El Hassouny et al. (2024) [7]	Unclear(5)	Low	Low	High(6)	High	Unclear(5)	Low	Low	Unclear

Author (year)	Risk of Bias					Applicability			
	Participants	Predictors	Outcome	Analysis	Overall	Participants	Predictors	Outcome	Overall
Ibrahim et al. (2022) [8]	Low	Low	High(7)	High(3)	High	Low	Low	High(8)	High
Kessler et al. (2023) [9]	Low	Low	High(1)	High(9)	High	Low	Low	High(1)	High
Khodadadi et al. (2023) [10]	Low	Low	Low	Unclear(2)	Unclear	Low	Low	Low	Low
Lim et al. (2025) [11]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High
Lin et al. (2019) [12]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High
Lu et al. (2020) [13]	Low	Low	Unclear(4)	Unclear(2, 10)	Unclear	Low	Low	Low	Low
Lu et al. (2021) [14]	Unclear(11)	Low	Unclear(4)	Unclear(11)	Unclear	Low	Low	Low	Low
Moazemi et al. (2022) [15]	Low	Low	Low	High(3)	High	Low	Low	Low	Low
Niu et al. (2023) [16]	Unclear(5)	Low	High(7)	Unclear(2,5,10,12)	High	Unclear(5)	Low	Low	Unclear
Pei et al. (2021) [17]	Unclear(5)	Low	Low	Unclear(2,5,10)	Unclear	Unclear(5)	Low	Low	Unclear
Pishgar et al. (2022) [18]	Low	Low	Low	High(3)	High	Low	Low	Low	Low
Rasheed et al. (2024) [19]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High

Author (year)	Risk of Bias					Applicability			
	Participants	Predictors	Outcome	Analysis	Overall	Participants	Predictors	Outcome	Overall
Sheetrit et al. (2023) [20]	Low	Low	Low	Low	Low	Low	Low	Low	Low
Shickel et al. (2022) [21]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High
Sun et al. (2024) [22]	Unclear(11)	Low	Low	High(9)	High	Low	Low	Low	Low
Wang et al. (2023) [23]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High
Zebin and Chaussalet (2019) [24]	Low	Low	High(1)	Low	High	Low	Low	High(1)	High

Notes

(1) Composite outcome including mortality

(2) No information about continuous and categorical predictors handling

(3) ICU/diagnosis subpopulation

(4) No information about number of events

(5) No information about inclusion and exclusion criteria

(6) No performance metric reported

(7) Different readmission time intervals considered

(8) Model performance was averaged across different readmission time intervals

(9) Listwise deletion

(10) No information about missing data handling

(11) No information about number of participants

(12) No information about internal validation

Supplementary Table 5. Summary of the comparison with statistical and/or traditional machine learning approach. The table summarizes, for each study that included a comparison with a statistical and/or traditional machine learning approach, the (sub)population, the timeframe of the ICU readmission outcome, and the statistical or traditional machine learning approach(es) used. For each model (defined as a combination of subpopulation, timeframe, and approach), the table reports the performance obtained during internal and external validation (when performed, as indicated by the notes) along with the corresponding 95% confidence interval (CI) or standard error (SE), when available.

Author (year)	**Population**	**Readmission timeframe**	**Model**	**Performance metric**	**Value**	**95% CI/SE**
Barbieri et al. (2020) [2]	All ICUs/diagnoses	30 days	LR	AUROC	0.659	0.656 - 0.663
				Precision	0.257	0.248 - 0.266
				F1 Score	0.296	0.291 - 0.300
				Sensitivity	0.606	0.597 - 0.615
				Specificity	0.647	0.639 - 0.655
Cao et al. (2023) [3]	Medical ICU(1)	Within hospital stay	SVM	Accuracy	0.7690	0.004
				PRAUC	0.2310	0.004
		Within hospital stay	KNN	Accuracy	0.6963	0.003

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	Medical ICU(1)			PRAUC	0.2306	0.004
	Medical ICU(2)	Within hospital stay	SVM	Accuracy	0.8882	0.002
				PRAUC	0.1116	0.002
	Medical ICU(2)	Within hospital stay	KNN	Accuracy	0.8637	0.003
				PRAUC	0.1113	0.002
Chiu et al. (2024) [4]	All ICUs/diagnoses	30 days	SVM	AUROC	0.6987	0.6945 – 0.7029
				Precision	0.2485	0.0032
				F1 Score	0.3552	0.0041
				Accuracy	0.7579	0.0027
				Recall	0.6221	0.0049
	All ICUs/diagnoses	30 days	RF	AUROC	0.6614	0.6536 – 0.669)
				Precision	0.8656	0.0149
				F1 Score	0.4592	0.0154
				Accuracy	0.9171	0.0012
				Recall	0.3126	0.0124

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	30 days	Gradient Boosting	AUROC	0.7193	0.7151 – 0.7235
				Precision	0.2778	0.0046
				F1 Score	0.3872	0.0047
				Accuracy	0.7826	0.0035
				Recall	0.6387	0.0053
	All ICUs/diagnoses	30 days	XGBoost	AUROC	0.6921	0.6899 – 0.6943
				Precision	0.4204	0.0100
				F1 Score	0.4396	0.0048
				Accuracy	0.8736	0.0022
				Recall	0.4609	0.0042
	All ICUs/diagnoses	30 days	LightGBM	AUROC	0.7205	0.7185 – 0.7225
				Precision	0.3632	0.0078
				F1 Score	0.4395	0.0052
				Accuracy	0.8473	0.0033
				Recall	0.5568	0.0017

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	30 days	CatBoost	AUROC	0.7065	0.7003 – 0.7127
				Precision	0.4468	0.0087
				F1 Score	0.4652	0.0038
				Accuracy	0.8795	0.0015
				Recall	0.4855	0.0116
Kessler et al. (2023) [9]	CSRU/CCU	2 days	LR	AUROC	0.708	-
				Precision	0.177	-
				F1 Score	0.274	-
				PRAUC	0.430	-
				Recall	0.597	-
	CSRU/CCU	2 days	RF	AUROC	0.777	-
				Precision	0.340	-
				F1 Score	0.312	-
				PRAUC	0.269	-
				Recall	0.288	-

Author (year)	**Population**	**Readmission timeframe**	**Model**	**Performance metric**	**Value**	**95% CI/SE**
	CSRU/CCU	2 days	ET	AUROC	0.776	-
				Precision	0.261	-
				F1 Score	0.290	-
				PRAUC	0.282	-
				Recall	0.513	-
	CSRU/CCU	2 days	FNN	AUROC	0.640	-
				Precision	0.312	-
				F1 Score	0.305	-
				PRAUC	0.281	-
				Recall	0.299	-
Khodadadi et al. (2023) [10]	All ICUs/diagnoses	Within hospital stay	RF	AUROC	0.4528	-
				PRAUC	0.8136	-
	All ICUs/diagnoses	Within hospital stay	MLP	AUROC	0.4519	-
				PRAUC	0.8121	-

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	Within hospital stay	LR	AUROC	0.4628	-
				PRAUC	0.8192	-
	All ICUs/diagnoses	Within hospital stay	SVM	AUROC	0.4612	-
				PRAUC	0.8213	-
	All ICUs/diagnoses	Within hospital stay	NB	AUROC	0.3491	-
				PRAUC	0.7605	-
	All ICUs/diagnoses	Within hospital stay	KNN	AUROC	0.4016	-
				PRAUC	0.8001	-
	All ICUs/diagnoses	Within hospital stay	XGBoost	AUROC	0.4481	-
				PRAUC	0.8083	-
Lim et al. (2025) [11]	All ICUs/diagnoses	Within hospital stay	MEWS	AUROC	0.631	0.615 – 0.647
				AUROC[3]	0.586	0.563 – 0.608
				AUROC[4]	0.601	0.586 – 0.615
	All ICUs/diagnoses		MEWS	AUROC	0.623	0.590 – 0.655

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
		2 days (within hospital stay)		AUROC(3)	0.608	0.565 – 0.650
				AUROC(4)	0.611	0.591 – 0.632
	All ICUs/diagnoses	More than 2 days (within hospital stay)	MEWS	AUROC	0.633	0.616 – 0.651
				AUROC(3)	0.577	0.553 – 0.602
				AUROC(4)	0.591	0.572 – 0.610
	All ICUs/diagnoses	Within hospital stay	NEWS	AUROC	0.661	0.644 – 0.677
				AUROC(3)	0.596	0.575 – 0.617
				AUROC(4)	0.632	0.618 – 0.646
	All ICUs/diagnoses	2 days (within hospital stay)	NEWS	AUROC	0.652	0.619 – 0.685
				AUROC(3)	0.622	0.581 – 0.663
				AUROC(4)	0.636	0.616 – 0.656
	All ICUs/diagnoses	More than 2 days (within hospital stay)	NEWS	AUROC	0.662	0.645 – 0.680
				AUROC(3)	0.586	0.563 – 0.610
				AUROC(4)	0.627	0.609 – 0.646

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	Within hospital stay	SWIFT	AUROC	0.680	0.664 – 0.695
				AUROC(3)	0.672	0.651 – 0.694
				AUROC(4)	0.656	0.642 – 0.671
	All ICUs/diagnoses	2 days (within hospital stay)	SWIFT	AUROC	0.607	0.576 – 0.638
				AUROC(3)	0.646	0.603 – 0.689
				AUROC(4)	0.646	0.625 – 0.666
	All ICUs/diagnoses	More than 2 days (within hospital stay)	SWIFT	AUROC	0.700	0.682 – 0.717
				AUROC(3)	0.680	0.657 – 0.703
				AUROC(4)	0.666	0.648 – 0.684
	All ICUs/diagnoses	Within hospital stay	APACHE-II	AUROC	0.632	0.615 – 0.648
				AUROC(3)	0.685	0.665 – 0.706
				AUROC(4)	0.684	0.670 – 0.697
	All ICUs/diagnoses	2 days (within hospital stay)	APACHE-II	AUROC	0.589	0.554 – 0.625
				AUROC(3)	0.700	0.662 – 0.738

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				AUROC(4)	0.676	0.657 – 0.695
	All ICUs/diagnoses	More than 2 days (within hospital stay)	APACHE-II	AUROC	0.642	0.624 – 0.660
				AUROC(3)	0.679	0.656 – 0.702
				AUROC(4)	0.691	0.674 – 0.708
	All ICUs/diagnoses	Within hospital stay	LR	AUROC	0.722	0.706 – 0.738
				AUROC(3)	0.709	0.687 – 0.732
				AUROC(4)	0.678	0.664 – 0.692
	All ICUs/diagnoses	2 days (within hospital stay)	LR	AUROC	0.656	0.622 – 0.689
				AUROC(3)	0.696	0.654 – 0.739
				AUROC(4)	0.650	0.629 – 0.672
	All ICUs/diagnoses	More than 2 days (within hospital stay)	LR	AUROC	0.740	0.721 – 0.758
				AUROC(3)	0.714	0.689 – 0.740
				AUROC(4)	0.703	0.683 – 0.723
	All ICUs/diagnoses	Within hospital stay	RF	AUROC	0.792	0.779 – 0.806

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				AUROC(3)	0.627	0.604 – 0.651
				AUROC(4)	0.636	0.621 – 0.650
	All ICUs/diagnoses	2 days (within hospital stay)	RF	AUROC	0.739	0.710 – 0.769
				AUROC(3)	0.607	0.563 – 0.650
				AUROC(4)	0.626	0.605 – 0.648
	All ICUs/diagnoses	More than 2 days (within hospital stay)	RF	AUROC	0.806	0.792 – 0.821
				AUROC(3)	0.634	0.608 – 0.660
				AUROC(4)	0.643	0.624 – 0.663
	All ICUs/diagnoses	Within hospital stay	Cox	AUROC	0.692	0.677 – 0.708
				AUROC(3)	0.679	0.658 – 0.700
				AUROC(4)	0.630	0.616 – 0.644
	All ICUs/diagnoses	2 days (within hospital stay)	Cox	AUROC	0.665	0.633 – 0.697
				AUROC(3)	0.664	0.625 – 0.703
				AUROC(4)	0.606	0.584 – 0.628

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
	All ICUs/diagnoses	More than 2 days (within hospital stay)	Cox	AUROC	0.699	0.682 – 0.716
				AUROC(3)	0.684	0.660 – 0.707
				AUROC(4)	0.651	0.631 – 0.670
Lin et al. (2019) [12]	All ICUs/diagnoses	30 days	LR	AUROC	0.777	0.765 – 0.789
				Recall	0.680	0.662 – 0.697
	All ICUs/diagnoses	30 days	NB	AUROC	0.709	0.702 – 0.716
				Recall	0.453	0.434 – 0.472
	All ICUs/diagnoses	30 days	RF	AUROC	0.714	0.703 – 0.725
				Recall	0.563	0.548 – 0.578
	All ICUs/diagnoses	30 days	SVM	AUROC	0.779	0.768 – 0.789
				Recall	0.703	0.685 – 0.720
Pei et al. (2021) [17]	All ICUs/diagnoses	30 days	SVM	AUROC	0.655	0.651 - 0.658
				Precision	0.265	0.256 - 0.274
				F1 Score	0.303	0.297 - 0.309
				Sensitivity	0.565	0.552 - 0.577

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				Specificity	0.679	0.668 - 0.691
Pishgar et al. (2022) [18]	Heart failure	30 days	RF	AUROC	0.713	0.691 – 0.761
				Precision	0.750	-
				F1 Score	0.760	-
				Sensitivity	0.801	-
				Accuracy	0.828	-
	Heart failure	30 days	XGBoost	AUROC	0.701	0.685 - 0.756
				Precision	0.731	-
				F1 Score	0.763	-
				Sensitivity	0.804	-
				Accuracy	0.826	-
	Heart failure	30 days	CatBoost	AUROC	0.704	0.674 – 0.759
				Precision	0.692	-
				F1 Score	0.752	-
				Sensitivity	0.800	-

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				Accuracy	0.829	-
	Heart failure	30 days	SVM	AUROC	0.680	0.657 – 0.712
				Precision	0.691	-
				F1 Score	0.753	-
				Sensitivity	0.801	-
				Accuracy	0.828	-
	Heart failure	30 days	Decision Tree	AUROC	0.689	0.669 – 0.721
				Precision	0.724	-
				F1 Score	0.710	-
				Sensitivity	0.691	-
				Accuracy	0.688	-
	Heart failure	30 days	KNN	AUROC	0.696	0.657 – 0.731
				Precision	0.725	-
				F1 Score	0.751	-
				Sensitivity	0.802	-

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				Accuracy	0.815	-
Rasheed et al. (2024) [19]	All ICUs/diagnoses	30 days	SVM	AUROC	0.789	-
				Precision	0.722	-
				Accuracy	0.719	-
				PRAUC	0.776	-
				Recall	0.713	-
	All ICUs/diagnoses	30 days	LR	AUROC	0.774	-
				Precision	0.718	-
				Accuracy	0.702	-
				PRAUC	0.765	-
				Recall	0.666	-
	All ICUs/diagnoses	30 days	NB	AUROC	0.714	-
				Precision	0.708	-
				Accuracy	0.642	-
				PRAUC	0.733	-

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				Recall	0.483	-
	All ICUs/diagnoses	30 days	KNN	AUROC	0.756	-
				Precision	0.680	-
				Accuracy	0.687	-
				PRAUC	0.741	-
				Recall	0.706	-
Shickel et al. (2022) [21]	All ICUs/diagnoses	Within hospital stay	CatBoost	AUROC	0.759	-
	All ICUs/diagnoses	Within hospital stay	XGBoost	AUROC	0.762	-
Sun et al. (2024) [22]	All ICUs/diagnoses	Within hospital stay	LR	AUROC	0.8081	0.0093
				F1 Score	0.7515	0.0106
				PRAUC	0.8433	0.0099
	All ICUs/diagnoses	Within hospital stay	RF	AUROC	0.8222	0.0080
				F1 Score	0.7711	0.0089
				PRAUC	0.8553	0.0061
Wang et al. (2023) [23]	All ICUs/diagnoses	30 days	LR	AUROC	0.7493	0.7328 - 0.7658

Author (year)	Population	Readmission timeframe	Model	Performance metric	Value	95% CI/SE
				Precision	0.8724	0.8697 - 0.8751
				F1 Score	0.4822	0.4681 - 0.4964
				Accuracy	0.7059	0.6981 - 0.7138
				PRAUC	0.4509	0.4153 - 0.4865
	All ICUs/diagnoses	30 days	RF	AUROC	0.7874	0.7853 - 0.7895
				Precision	0.7863	0.7684 - 0.8041
				F1 Score	0.2481	0.2428 - 0.2533
				Accuracy	0.8126	0.8116 - 0.8137
				PRAUC	0.5355	0.5315 - 0.5394
	All ICUs/diagnoses	30 days	XGBoost	AUROC	0.7493	0.7328 - 0.7658
				Precision	0.6417	0.6272 - 0.6561
				F1 Score	0.4257	0.4178 - 0.4336
				Accuracy	0.8199	0.8172 - 0.8226
				PRAUC	0.5488	0.5433 - 0.5542
Zebin and Chaussalet (2019) [24]	All ICUs/diagnoses	30 days	LR	AUROC	0.714	-

Author (year)	**Population**	**Readmission timeframe**	**Model**	**Performance metric**	**Value**	**95% CI/SE**
				Precision	0.872	-
				Accuracy	0.703	-
				Recall	0.736	-
	All ICUs/diagnoses	30 days	RF	AUROC	0.770	-
				Precision	0.892	-
				Accuracy	0.723	-
				Recall	0.756	-
	All ICUs/diagnoses	30 days	SVM	AUROC	0.775	-
				Precision	0.904	-
				Accuracy	0.711	-
				Recall	0.722	-

Notes

(1) Internal validation on MIMIC-III

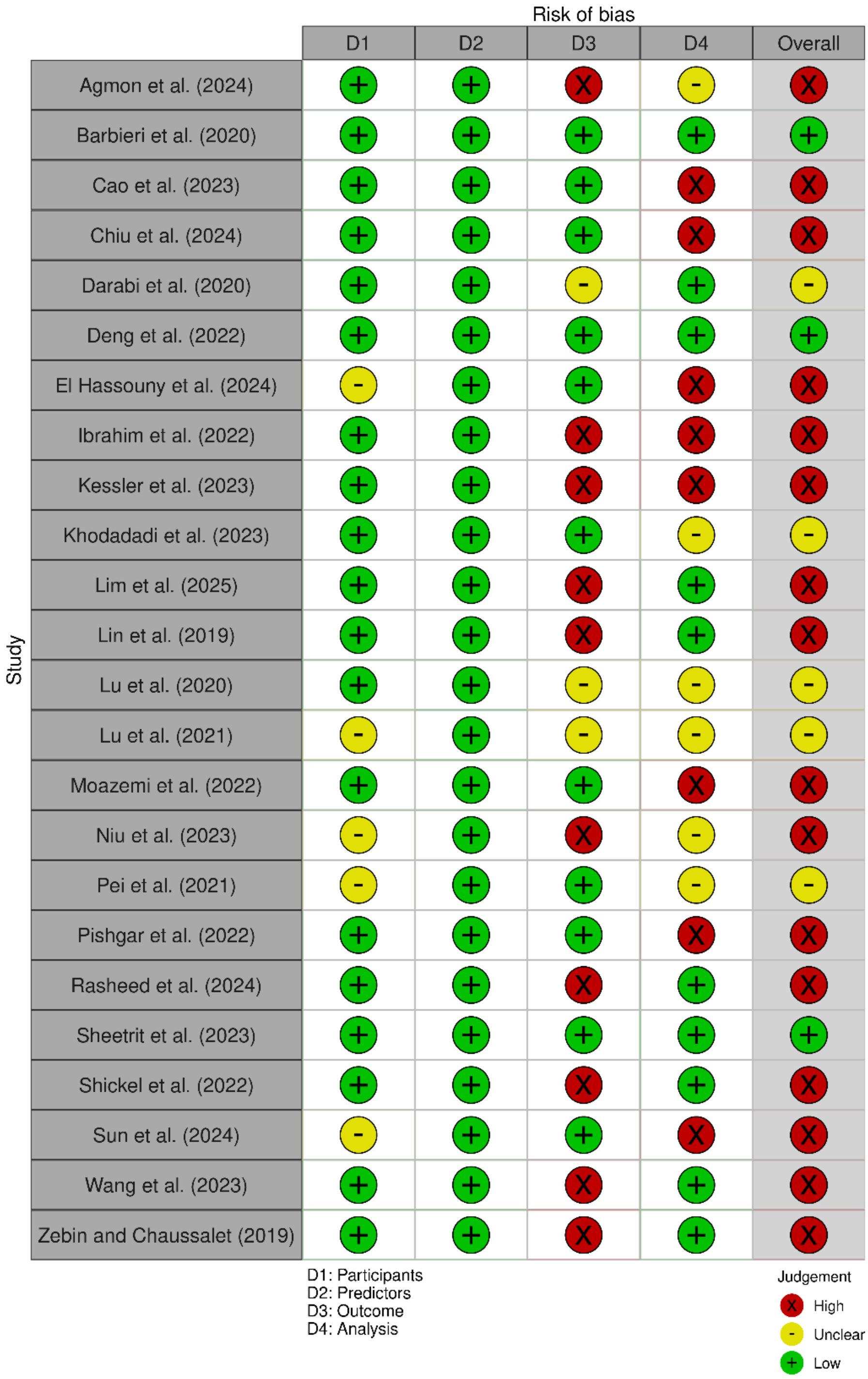
(2) Internal validation on eICU

(3) External validation on MIMIC-III

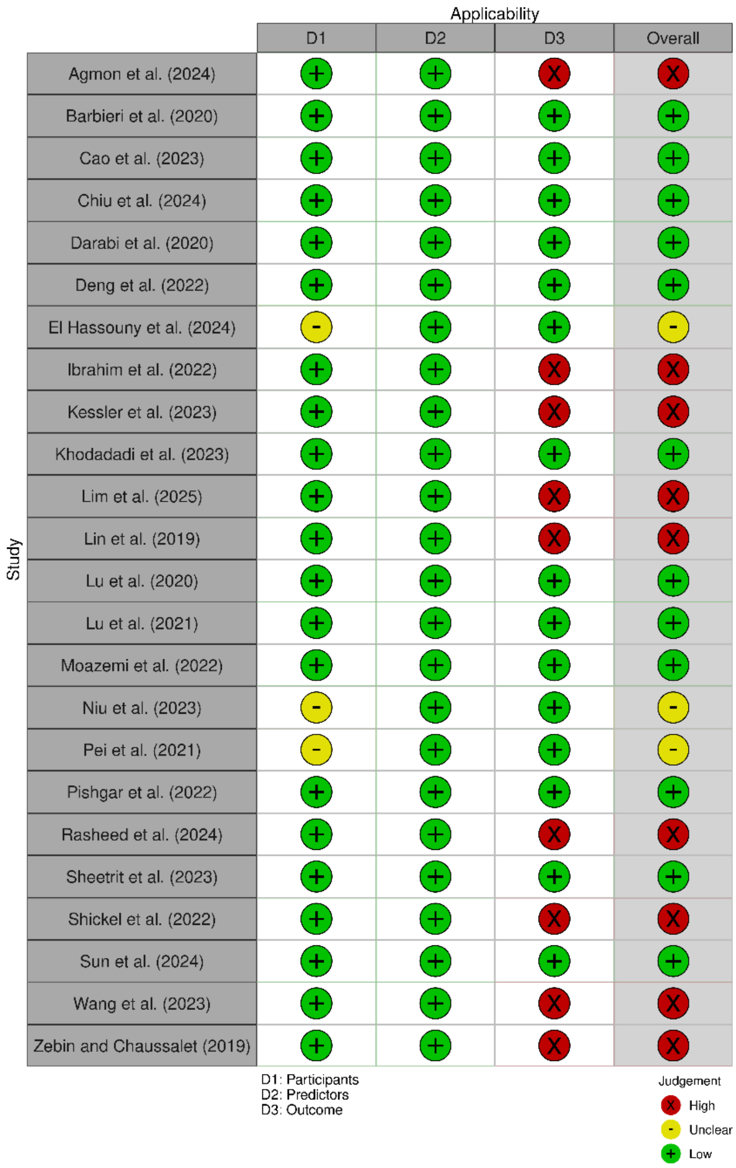
(4) External validation on eICU

Abbreviations: APACHE-II, acute physiology and chronic health evaluation II; AUROC, area under the receiver operating characteristic; CI, confidence interval; Cox, Cox proportional hazards regression; CSRU, cardiosurgery recovery unit; CCU, coronary care unit; ET, extra trees; FNN, feedforward neural network; KNN, k-nearest neighbors; LR, logistic regression; MEWS, modified early warning score; MLP, multilayer perceptron; NB, naive Bayes; NEWS, national early warning score; PRAUC, area under the precision-recall curve; RF, random forests; SVM, support vector machine; SWIFT, stability and workload index for transfer.

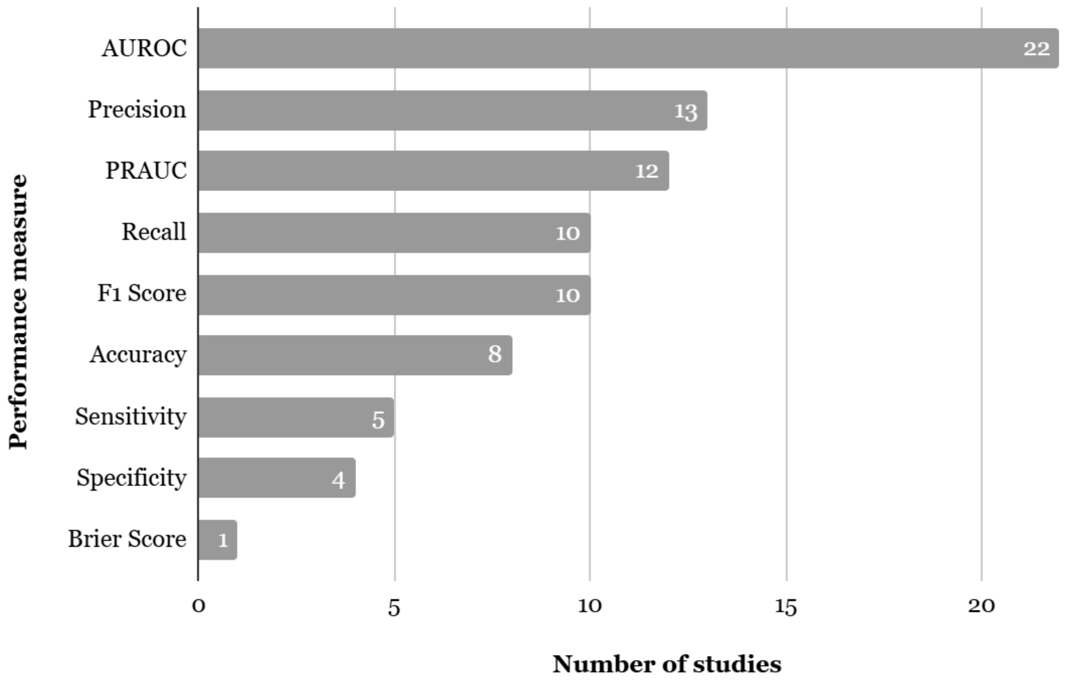
Supplementary Figures



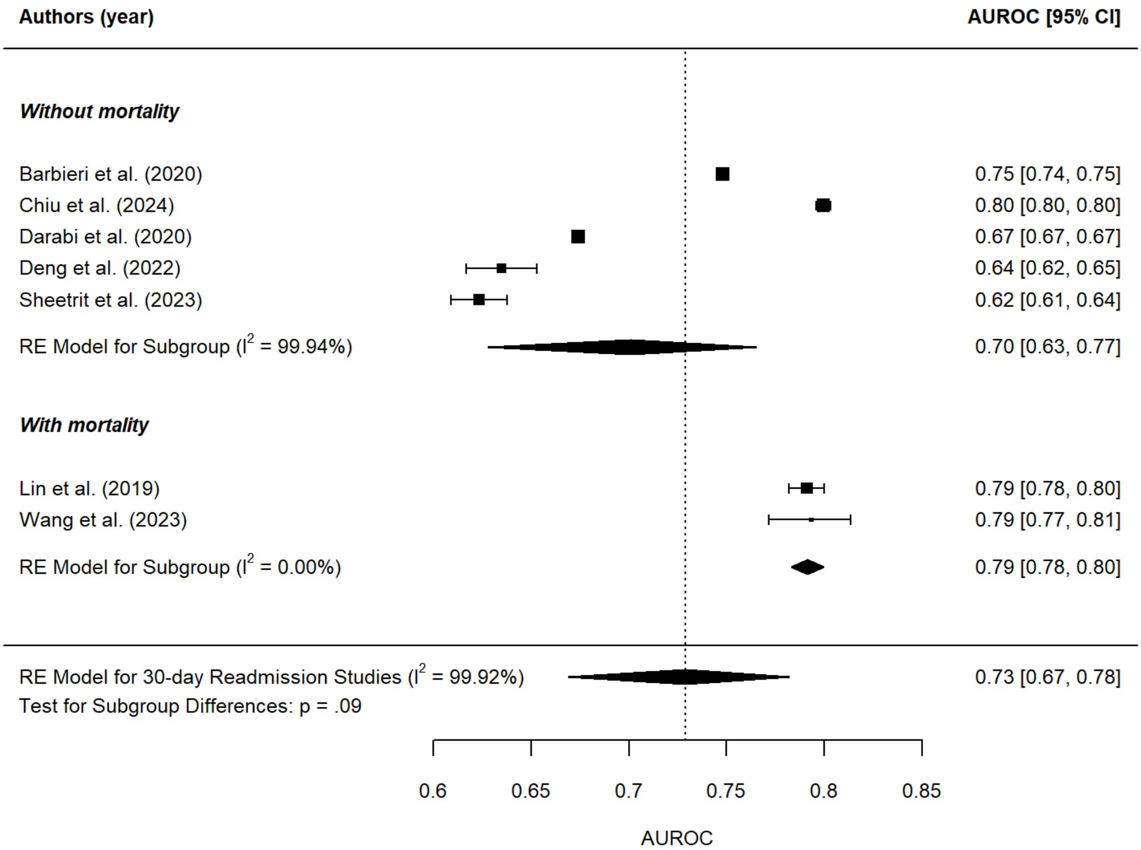
Supplementary Figure 1. Traffic light plot of PROBAST assessment (risk of bias domains). A green circle with a plus sign indicates low risk of bias; a yellow circle with a minus sign indicates unclear risk of bias; and a red circle with a cross sign indicates high risk of bias.



Supplementary Figure 2. Traffic light plot of PROBAST assessment (Applicability domains). A green circle with a plus sign indicates low risk of bias; a yellow circle with a minus sign indicates unclear risk of bias; and a red circle with a cross sign indicates high risk of bias.

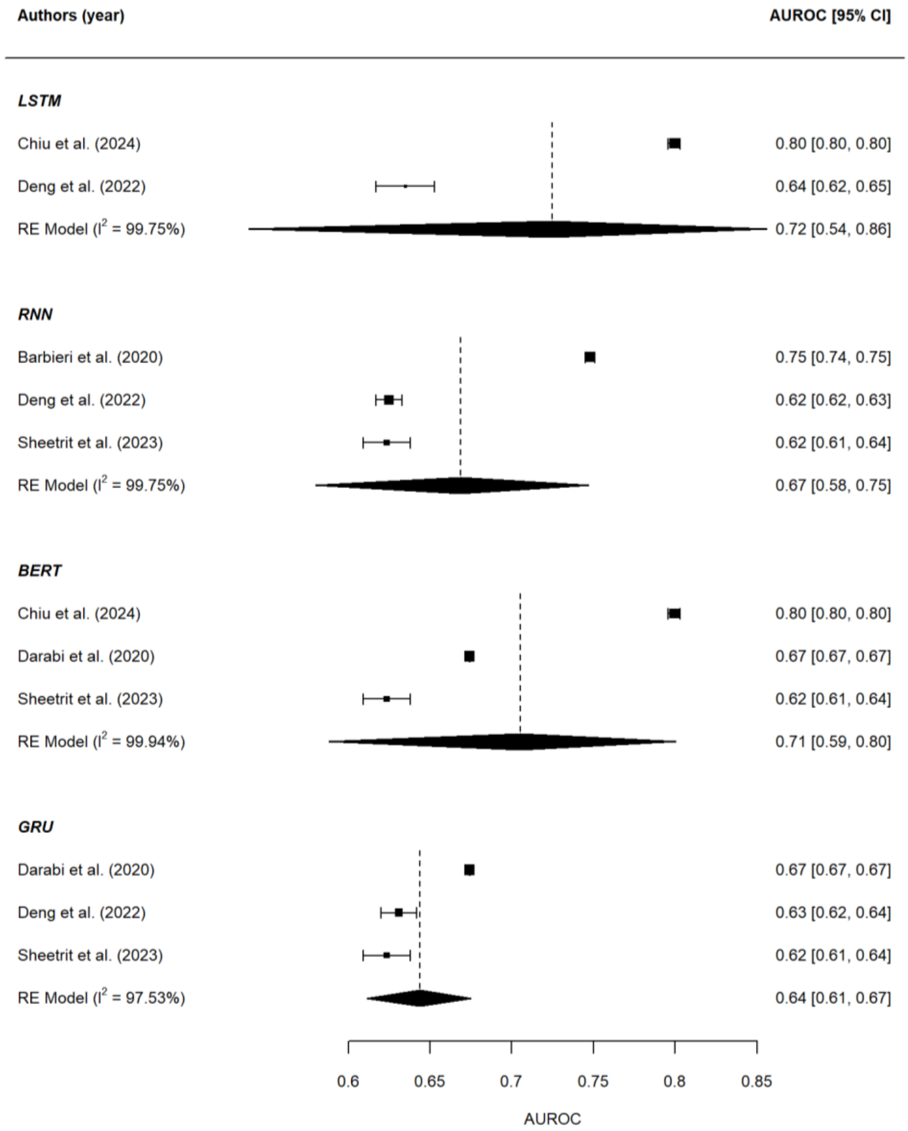


Supplementary Figure 3. Model performance metrics reported across the 24 included studies. The bar plot shows the absolute frequencies with which each metric was reported. Several studies report more than one metric, as detailed in Supplementary Table 3.



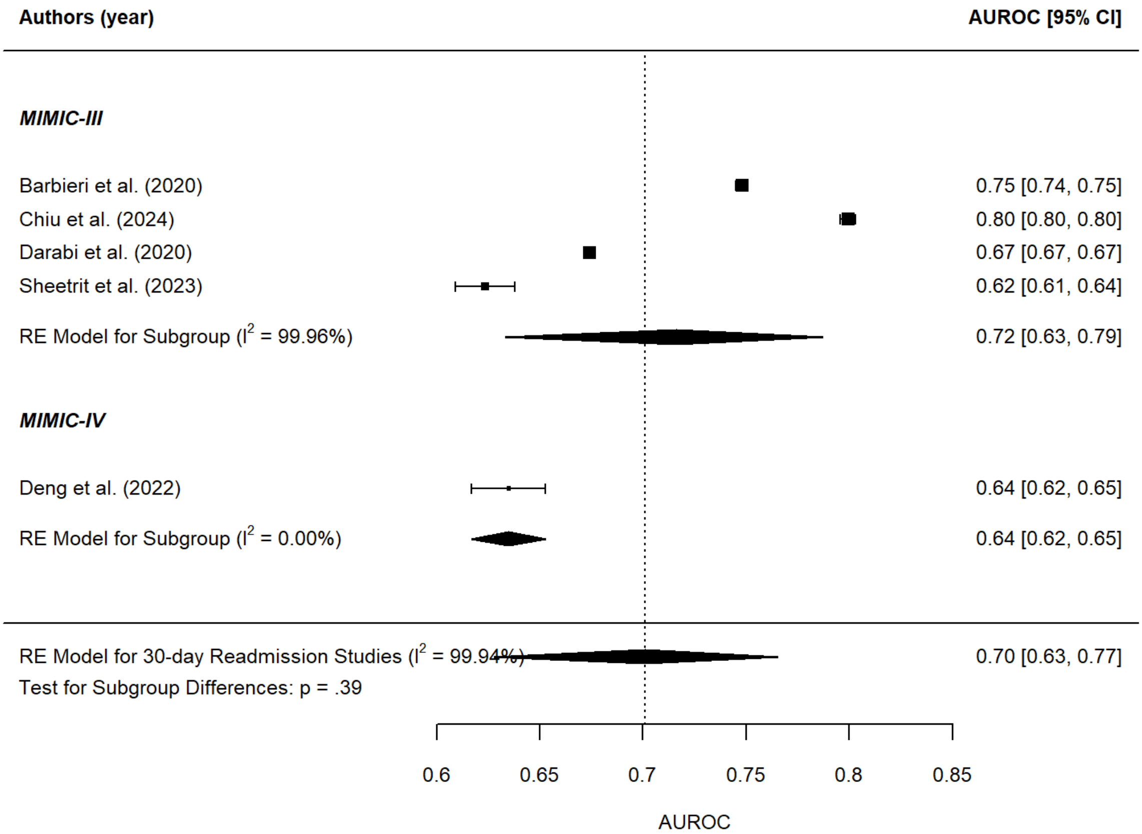
Supplementary Figure 4. Forest plot for the meta-analysis of studies predicting 30-day readmission including and excluding mortality as a composite outcome. The plot shows the meta-analytic receiver operating characteristic (AUROC) metric resulted from a subgroup analysis based on whether mortality was included as part of the composite outcome. The 95% confidence intervals (CI) and I^2 statistic are reported along with the estimates. The p-value (p) for the statistical test of subgroup differences is also reported. We excluded studies not defining the outcome as 30-day ICU readmission, and those focused on a disease-specific subpopulation. For each study, we selected the best-performing model.

Abbreviations: RE, random effects.



Supplementary Figure 5. Forest plot for the meta-analysis of different deep learning architecture. The plot shows the meta-analytic receiver operating characteristic (AUROC) metric resulted from a subgroup analysis by deep learning architecture. The 95% confidence intervals (CI) and I^2 statistic are reported along with the estimates. The p-value (p) for the statistical test of subgroup differences is also reported. We excluded studies not defining the outcome as 30-day ICU readmission excluding mortality, and those focused on a disease-specific subpopulation. For each study, we selected the best-performing model for each architecture.

Abbreviations: RE, random effects.



Supplementary Figure 6. Forest plot for the meta-analysis of different data sources (MIMIC-III and IV). The plot shows the meta-analytic receiver operating characteristic (AUROC) metric resulting from a subgroup analysis by dataset used (MIMIC-III or MIMIC-IV). The 95% confidence intervals (CI) and I^2 statistic are reported along with the estimates. The p-value (p) for the statistical test of subgroup differences is also reported. We excluded studies not defining the outcome as 30-day ICU readmission excluding mortality, and those focused on a disease-specific subpopulation For each study, we selected the best-performing model.

Abbreviations: RE, random effects.

References

1. Agmon S, Singer U, Radinsky K. Leveraging Temporal Trends for Training Contextual Word Embeddings to Address Bias in Biomedical Applications: Development Study. JMIR AI. 2024;3:e49546.
2. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk. Sci Rep. 2020;10:1111.
3. Cao Y, Wang Q, Wang X, Peng D, Li P. Multi-gate Mixture of Multi-view Graph Contrastive Learning on Electronic Health Record. IEEE J Biomed Health Inform. 2023;PP.
4. Chiu C-C, Wu C-M, Chien T-N, Kao L-J, Li C. Predicting ICU Readmission from Electronic Health Records via BERTopic with Long Short Term Memory Network Approach. J Clin Med. 2024;13:5503.
5. Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. TAPER: Time-Aware Patient EHR Representation. IEEE Journal of Biomedical and Health Informatics. 2020;24:3268–75.
6. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. Front Med (Lausanne). 2022;9:933037.
7. El Hassouny A, El Bouanani F, Qaraqe KA. Mathematical Modeling and Optimal Stopping Theory-Based Extra Layers for 30-Day Rate Risk Prediction of Readmission to Intensive Care Units. IEEE Transactions on Artificial Intelligence. 2024;5:2723–38.
8. Ibrahim ZM, Bean D, Searle T, Qian L, Wu H, Shek A, et al. A Knowledge Distillation Ensemble Framework for Predicting Short- and Long-Term Hospitalization Outcomes From Electronic Health Records Data. IEEE J Biomed Health Inform. 2022;26:423–35.

9. Kessler S, Schroeder D, Korlakov S, Hettlich V, Kalkhoff S, Moazemi S, et al. Predicting readmission to the cardiovascular intensive care unit using recurrent neural networks. Digit Health. 2023;9:20552076221149529.
10. Khodadadi A, Ghanbari Bousejin N, Molaei S, Kumar Chauhan V, Zhu T, Clifton DA. Improving Diagnostics with Deep Forest Applied to Electronic Health Records. Sensors. 2023;23.
11. Lim L, Kim M, Cho K, Yoo D, Sim D, Ryu HG, et al. Multicenter validation of a machine learning model to predict intensive care unit readmission within 48 hours after discharge. eClinicalMedicine [Internet]. 2025 [cited 2025 Feb 24];81. Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(25)00044-6/fulltext
12. Lin Y-W, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One. 2019;14:e0218942.
13. Lu Q, de Silva N, Dou D, Nguyen TH, Sen P, Reinwald B, et al. Exploiting Node Content for Multiview Graph Convolutional Network and Adversarial Regularization. 2020. p. 545–55.
14. Lu Q, Nguyen TH, Dou D. Predicting Patient Readmission Risk from Medical Text via Knowledge Graph Enhanced Multiview Graph Convolution. 2021. p. 1990–4.
15. Moazemi S, Kalkhoff S, Kessler S, Boztoprak Z, Hettlich V, Liebrecht A, et al. Evaluating a Recurrent Neural Network Model for Predicting Readmission to Cardiovascular ICUs Based on Clinical Time Series Data †. Engineering Proceedings. 2022;18.
16. Niu K, Pei S, Peng X, Zeng J, Zhang K. Intensive Care Unit readmission prediction with correlation enhanced multi-task learning. Computers and Electrical Engineering. 2023;110:108780.

17. Pei S, Niu K, Peng X, Zeng J. Readmission Prediction with Knowledge Graph Attention and RNN-Based Ordinary Differential Equations. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2021;12817 LNAI:559–70.
18. Pishgar M, Theis J, Del Rios M, Ardati A, Anahideh H, Darabi H. Prediction of unplanned 30-day readmission for ICU patients with heart failure. BMC Med Inform Decis Mak. 2022;22:117.
19. Rasheed N, Akram MU, Zeb J. Convolutional Neural Network Based Predictive Framework for Intensive Care Unit Readmission. 2024 5th International Conference on Advancements in Computational Sciences (ICACS) [Internet]. 2024 [cited 2025 Mar 10]. p. 1–9. Available from: https://ieeexplore.ieee.org/document/10473298/?arnumber=10473298
20. Sheetrit E, Brief M, Elisha O. Predicting unplanned readmissions in the intensive care unit: a multimodality evaluation. Sci Rep. 2023;13:15426.
21. Shickel B, Silva B, Ozrazgat-Baslanti T, Ren Y, Khezeli K, Guan Z, et al. Multi-dimensional patient acuity estimation with longitudinal EHR tokenization and flexible transformer networks. Front Digit Health. 2022;4:1029191.
22. Sun M, Yang X, Niu J, Gu Y, Wang C, Zhang W. A cross-modal clinical prediction system for intensive care unit patient outcome. Knowledge-Based Systems. 2024;283:111160.
23. Wang L, Zhao L, Luo Z, Wang X, Feng Z, Lu L. CARE-30: A Causally Driven Multi-Modal Model for Enhanced 30-Day ICU Readmission Predictions. 2023. p. 1509–16.
24. Zebin T, Chaussalet TJ. Design and implementation of a deep recurrent model for prediction of readmission in urgent care using electronic health records. 2019.