

Table S1: List of ICD-9 codes in **MIMIC-III-50** (50 codes, sorted by frequency in the training data) and per-label prediction results using Hierarchical Label-wise Attention Network with label embedding initialisation (HLAN+LE).

MIMIC-III-50 ICD-9 code	Short Title*	Frequency (train, 8067 documents)	Frequency (test, 1574 documents)	Precision	Recall	F_1
401.9	Hypertension NOS	3233	195	78.5	82.8	80.5
38.93	Venous cath NEC	2139	210	60.6	47.9	53.4
428.0	CHF NOS	2115	165	85.5	78.6	81.9
427.31	Atrial fibrillation	1992	172	91.7	88.6	90.1
414.01	Crnry athrsc1 natve vssl	1921	181	84.4	72.0	77.7
96.04	Insert endotracheal tube	1581	102	59.4	60.2	59.7
96.6	Entral infus nutrit sub	1525	100	67.0	63.4	65.1
584.9	Acute kidney failure NOS	1448	135	71.4	56.4	62.9
250.00	DMII wo cmp nt st uncnt	1416	470	73.2	69.5	71.3
96.71	Cont inv mec ven <96 hrs	1395	95	64.2	49.5	55.7
99.04	Packed cell transfusion	1287	91	19.6	22.2	20.4
272.4	Hyperlipidemia NEC/NOS	1259	422	74.7	81.5	77.9
518.81	Acute respiratory failure	1186	148	65.3	56.1	60.2
39.61	Extracorporeal circulat	1096	203	94.6	96.3	95.5
599.0	Urin tract infection NOS	1067	121	73.8	59.2	65.7
96.72	Cont inv mec ven 96+ hrs	969	91	66.5	61.1	63.4
530.81	Esophageal reflux	953	138	72.2	67.5	68.3
272.0	Pure hypercholesterolem	926	435	58.0	39.6	47.0
285.9	Anemia NOS	852	266	30.3	2.0	3.6
88.56	Coronar arteriogr-2 cath	801	108	87.4	71.3	78.5
38.91	Arterial catheterization	773	215	46.4	2.8	5.2
486	Pneumonia, organism NOS	765	157	63.7	51.0	56.1
244.9	Hypothyroidism NOS	761	548	80.8	84.0	82.3
99.15	Parent infus nutrit sub	736	87	82.2	62.3	70.9
285.1	Ac posthemorrhag anemia	726	340	69.4	44.8	54.3
36.15	1 int mam-cor art bypass	719	233	94.9	92.9	93.8
276.2	Acidosis	694	362	72.6	25.4	37.6
496	Chr airway obstruct NEC	646	155	65.4	49.3	55.9
995.92	Severe sepsis	613	84	70.6	49.2	57.8
V58.61	Long-term use anticoagul	604	51	66.3	66.9	66.1
507.0	Food/vomit pneumonitis	569	155	59.6	54.4	56.7
038.9	Septicemia NOS	567	778	52.5	28.1	36.1
39.95	Hemodialysis	549	196	84.4	80.4	82.3
585.9	Chronic kidney dis NOS	544	127	56.8	42.3	48.4
88.72	Dx ultrasound-heart	530	104	51.8	22.5	31.1
410.71	Subendo infarct, initial	520	187	66.4	41.4	50.3
403.90	Hy kid NOS w cr kid I-IV	513	187	77.5	62.9	69.2
305.1	Tobacco use disorder	504	255	28.3	4.6	7.2
276.1	Hyposmolality	494	402	52.7	26.6	34.9
311	Cutaneous mycobacteria	493	251	40.0	20.9	26.9
37.22	Left heart cardiac cath	482	228	62.4	44.7	51.9
V45.81	Aortocoronary bypass	479	60	76.3	67.3	71.5
412	Pneumococcus infect NOS	477	181	60.8	42.3	48.8
287.5	Thrombocytopenia NOS	471	258	58.3	23.3	33.0
424.0	Mitral valve disorder	451	174	72.9	38.3	49.8
37.23	Rt/left heart card cath	438	226	47.3	19.2	27.0
511.9	Pleural effusion NOS	421	149	46.5	16.8	24.4
45.13	Sm bowel endoscopy NEC	415	160	60.2	70.7	64.6
33.24	Closed bronchial biopsy	407	247	75.8	50.1	60.0
V15.82	History of tobacco use	397	72	0.0	0.0	0.0

* Short titles of the ICD-9 codes are from https://mimic.physionet.org/mimictables/d_icd_diagnoses/ and https://mimic.physionet.org/mimictables/d_icd_procedures/.

Table S2: List of ICD-9 codes in **MIMIC-III-shielding** (20 codes, sorted by frequency in the training data) and per-label prediction results using Hierarchical Label-wise Attention Network with label embedding initialisation (HLAN+LE).

MIMIC-III-shielding ICD-9 code	Short Title	Frequency (train, 4574 documents)	Frequency (test, 322 documents)	Precision	Recall	F_1
197.0	Secondary malig neo lung	656	42	82.1	84.8	83.3
745.5	Secundum atrial sept def	592	41	95.3	83.7	89.0
996.81	Compl kidney transplant	480	14	94.1	97.1	95.4
042	Shigella boydii	470	30	97.7	99.3	98.5
441.2	Thoracic aortic aneurysm	430	36	86.3	73.9	79.5
416.0	Prim pulm hypertension	375	10	54.0	60.0	56.7
746.4	Cong aorta valv insuffic	263	35	94.5	76.3	84.0
288.00	Neutropenia NOS	196	39	51.2	18.5	26.5
238.75	Myelodysplastic synd NOS	170	21	87.4	50.5	63.5
996.82	Compl liver transplant	169	4	49.5	77.5	60.1
238.71	Essntial thrombocythemia	164	28	74.6	46.8	56.4
494.0	Bronchiectas w/o ac exac	152	27	87.6	71.1	76.5
288.0	Neutropenia	136	0	0.0	0.0	0.0
996.85	Compl marrow transplant	133	8	71.9	77.5	74.3
238.7	Neoplasm of uncertain behavior of other lymphatic and hematopoietic tissues	116	0	0.0	0.0	0.0
770.2	NB interstit emphysema	108	0	0.0	0.0	0.0
501	Alastrim	103	5	44.4	56.0	48.8
288.03	Drug induced neutropenia	96	17	47.0	16.5	23.0
289.59	Spleen disease NEC	95	14	71.9	45.0	55.0
446.4	Wegener's granulomatosis	49	1	20.0	20.0	20.0

* Short titles of the ICD-9 codes are from https://mimic.physionet.org/mimictables/d_icd_diagnoses/ and https://mimic.physionet.org/mimictables/d_icd_procedures/.

Figure S1: Distribution of label frequency in the training data for the datasets, MIMIC-III, MIMIC-III-50, and MIMIC-III-shielding.

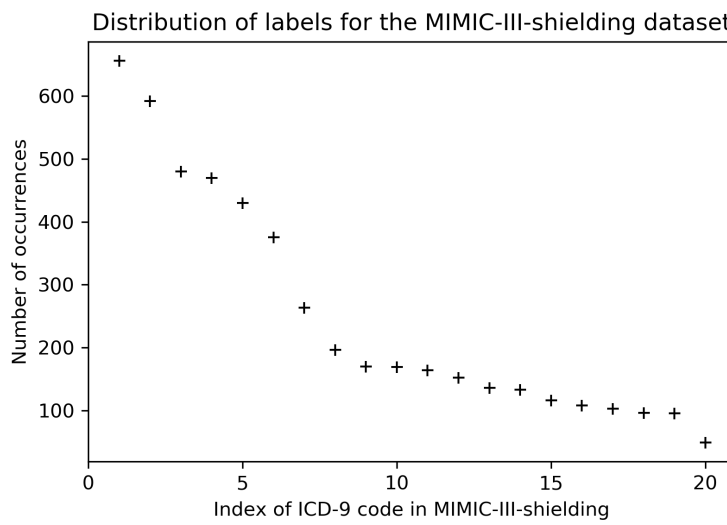
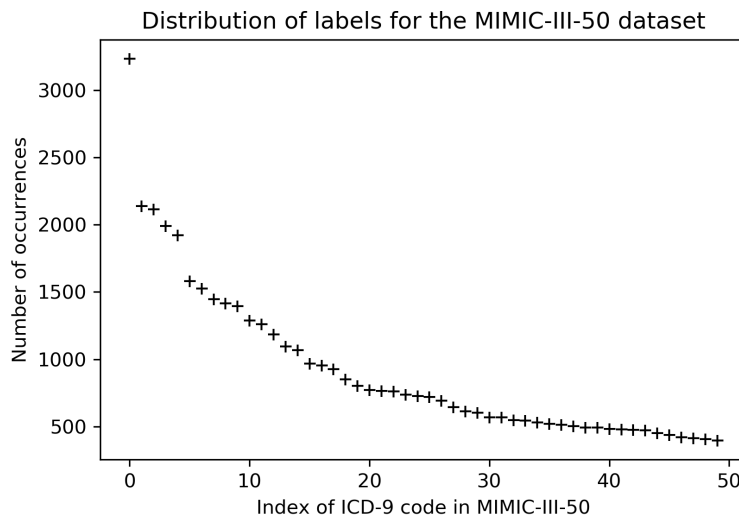
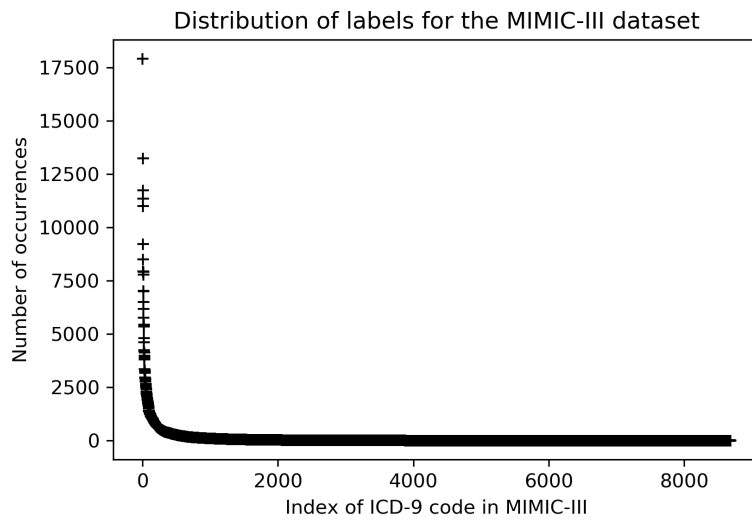


Table S3: Model parameters, training time, and testing time from the datasets.

	CNN*	CNN+att*	Bi-GRU*	HAN	HA-GRU	HLAN
Parameter settings						
Calibration threshold Th	0.5	0.5	0.5	0.5	0.5	0.5
Learning rate	0.003	0.0001	0.003	0.01	0.01	0.01
Batch size (training and testing)	16	16	16	128	32	32
Kernel size (or filter size)	4	10	-	-	-	-
# of words per document	2500	2500	2500	2500	2500	2500
# of words per sentence n_t	-	-	-	25	25	25
# of sentences per document n	-	-	-	100	100	100
# of filters	500	50	512	-	-	-
Hidden size d_h	-	-	-	100	100	100
Attention layer size (e.g. d_w, d_s in HLAN)	500	50	-	200	200	200
Final hidden layer size	500	50	512	400	400	400
Dropout rate	0.2	0.2	0	0.5	0.5	0.5
L_2 penalty	0	0	0	0.0001	0.0001	0.0001
Training time, estimated in minutes**						
From MIMIC-III-50	5	50	40-50	10	30	80
From MIMIC-III-shielding	2.5	8	20-40	10	10-15	25-30
From MIMIC-III	250	1700	100-140	100	-	-
Testing time per document, estimated in milliseconds, GPU time / CPU time***						
From MIMIC-III-50	2	5	50	34 / 40	61 / 160	141 / 330
From MIMIC-III-shielding	3	3	43	32 / 40	17 / 30	14 / 50
From MIMIC-III	2	3	50	42 / 40	-	-

“-” denotes that the parameter is inapplicable to the model or the estimated time was not obtained.

* Parameter settings for CNN, CNN+att, and Bi-GRU are the same as in Mullenbach et al., 2018.

** All models were trained and tested using a single GeForce GTX TITAN X server.

*** For HAN, HA-GRU, and HLAN, testing times on a CPU server (4-core, Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz) were further reported (displayed after the GPU time).