#### **BEHRT: Transformer for Electronic Health Records**

Mr. Yikuan Li, Mr. Shishir Rao, Dr. Jose Roberto Ayala Solares, Dr. Abdelaali Hassaine, Dr. Rema Ramakrishnan, Dr. Dexter Canoy, Dr. Yajie Zhu, Dr. Kazem Rahimi, Dr. Gholamreza Salimi-Khorshidi

### **Supplementary Information**

#### Legend:

Table S1:This table provides the basic statistics for the characteristics of the selected cohorts.

Table S2: This table shows the Hyperparameter tuning results for BEHRT. The columns are various hyperparameters and the rows are different runs. The bold items are best performing.

Table S3: This table shows the top ten associations for each disease that occurred in at least 1% of the population. Each row is a disease and the columns are the ten diseases associated to said disease. The annotation "Good" means that the medical validator defined the association as a medically valid association and "Not Good" means that the medical validator defined the association as medically invalid.

Table S4: This table shows the Hyperparameter tuning results for Deepr. The columns are various hyperparameters and the rows are different runs. The bold items are best performing.

Table S5: This table shows the Hyperparameter tuning results for RETAIN. The columns are various hyperparameters and the rows are different runs. The bold items are best performing.

Table S6: This table shows the individual, disease wise precision for BEHRT for the next 6 months prediction task. The ratio in column 5 means the occurrence of a particular disease code in the label of the task over all labels in test data.

Figure S1: This figure shows the graph for an understanding of APS and AUROC for the next 6 months prediction task for all three models denoted by colour of dot.

Table S7: This table shows the number of Male patient Predictions and Female patient Predictions with predictive probability above 0.5 for a gender-specific disease in the next 6 months prediction test dataset. Disease Gender represents whether a disease is a male or female specific disease and M and F implies Male and Female respectively.

Table S8: This table shows the performance of the ablation study conducted on the next 6 Months prediction task. By turning on and off various embeddings, we measure performance with Average Precision Score (APS) and Area Under the ROC Curve (AUROC).

# **Population Statistics**

Table S1: Statistics of Cohorts Selected Prediction Tasks

Charact	eristics	Next Visit	Next 6 M	Next 12M
Gender	Male	41.80%	42.30%	41.70%
	Female	58.20%	57.70%	58.30%
Ethnicity	White	46.40%	48.30%	47.40%
	Unknown	43.80%	44.00%	44.50%
	Indian	0.40%	0.50%	0.50%
	Other	0.30%	0.30%	0.30%
	Pakistani	0.20%	0.30%	0.20%
	Black Carib	0.20%	0.30%	0.20%
	Other Asian	0.10%	0.10%	0.10%
	Black African	0.10%	0.10%	0.10%
	Mixed	0.10%	0.10%	0.10%
	Bangladeshi	0.08%	0.07%	0.07%
	Black Other	0.07%	0.06%	0.06%
	Chinese	0.06%	0.06%	0.05%
Age Start	0.25 Quantile	45	46	46
	0.5 Quantile	58	60	59
	0.75 Quantile	70	71	70
Age End	0.25 Quantile	56	58	58
	0.5 Quantile	70	71	71
	0.75 Quantile	81	82	82
Unique Codes	0.25 Quantile	7	8	8
	0.5 Quantile	9	10	11
	0.75 Quantile	12	15	15
Number of Visits	0.25 Quantile	10	14	14
	0.5 Quantile	15	20	20
	0.75 Quantile	24	30	30

Age Start and Age End: Corresponding age for the first visit and last visit been recorded

# **Hyperparameter Tuning for MLM**

In this section, we show the hyperparameter tuning results. In Table S2, we show the model's performance during the process of training the MLM. We performed Bayesian Optimisation to search for optimal parameters for the model.

**Table S2.** MLM Hyperparameter Tuning

Iteration	Hidden Size	Layers	Attention Heads	Intermediate Size	Precision
1	216	3	6	256	0.6191
2	288	9	12	512	0.6399
3	216	3	12	512	0.6175
4	432	3	18	512	0.6397
5	288	6	6	784	0.6380
6	216	6	18	512	0.6262
7	288	3	18	512	0.6292
8	432	3	6	784	0.6426
9	288	6	12	512	0.6356
10	288	3	12	256	0.6283
11	432	9	18	512	0.6466
12	576	9	6	1024	0.6538
13	432	3	18	1024	0.6411
14	432	9	6	1024	0.6508
15	576	6	6	256	0.6503
16	576	6	12	256	0.6510
17	360	9	18	512	0.6404
18	576	9	6	512	0.6513
19	288	6	6	512	0.6363
20	288	3	6	512	0.6297
21	288	6	12	512	0.6597
22	576	3	12	512	0.6487
23	360	6	6	784	0.6412
24	432	9	6	512	0.6497
25	360	6	12	512	0.6423



## **Medical Evaluation of Associations**

Table S3 shows the medical evalution of our diseases associations based on highest consine similarity. Each row is a disease and the columns are the ten diseases associated to said disease. The annotation "Good" means that the medical validator defined the association as a medically valid association and "Not Good" means that the medical validator defined the association as medically invalid.

Table C2. Medical Validation	of Diagona Associations I	aaaad an Diaaaaa Enabaddinaa
Table 53: Medical Validation	of Disease Associations (	pased on Disease Embeddings

	Disease	1	2	3	4	5	6	7	8	9	10
											Endometrial
								Menorrhagia		Polycystic	hyperplasia
	Postcoital and	Postcoital and	Carcinoma in				Primary	and		ovarian	and
	contact	contact	situ cervical-	Dysmenorrhoe	Leiomyoma of	Endometriosis-	Malignancy	polymenorrho	Prematurity-	syndrome-	hypertrophy-
1	bleeding	bleeding-Good	Good	a-Not Good	uterus-Good	Good	Cervical-Good	ea-Not Good	Not Good	Good	Good
										Benign	
										neoplasm of	
							Diaphragmatic			stomach and	Supraventricul
		Cholecystitis-	Cholelithiasis-	Cholangitis-	Pancreatitis-	Appendicitis-	hernia-Not	Infection of	Fatty Liver-	duodenum-Not	ar tachycardia-
2	Cholecystitis	Good	Good	Good	Good	Not Good	Good	liver-Good	Good	Good	Not Good
											Secondary
								Lower			Malignancy
				Pulmonary			Atrioventricula	Respiratory	Pericardial		retroperitoneu
			Right bundle	collapse (excl	Urinary Tract	Atrioventricula	r block,	Tract	effusion		m and
	Acute Kidney	Acute Kidney	branch block-	pneumothorax	Infections-	r block, first	complete-	Infections-	(noninflammat	Trifascicular	peritoneum-
3	Injury	Injury-Good	Good	)-Good	Good	degree-Good	Good	Good	ory)-Good	block-Good	Good
					Endometrial						
				Benign	hyperplasia						Polycystic
	Benign	Benign		neoplasm and	and			Primary	Postmenopaus	Primary	ovarian
	neoplasm of	neoplasm of	Endometriosis-	polyp of uterus-	hypertrophy-	Leiomyoma of	Dysmenorrhoe	Malignancy	al bleeding-	Malignancy	syndrome-
4	ovary	ovary-Good	Good	Good	Good	uterus-Good	a-Good	Uterine-Good	Good	Cervical-Good	Good

5	Visual impairment and blindness	Visual impairment and blindness- Good	Macular degeneration- Good	Retinal vascular occlusions- Good	Retinal detachment- Good	Keratitis-Good	Posterior Uveitis-Good Peripheral neuropathies (excluding cranial nerve	Glaucoma- Good	Delirium, not induced by alcohol and other psychoactive substances- Not Good	Anterior and Intermediate Uveitis-Good	Cataract-Good
		Intervertebral					and carpal tunnel	Postinfective and reactive	Alopecia	Infection of bones and	Disorders of autonomic
ε	Intervertebral disc disorders	disc disorders- Good	Spondylolisthe sis-Good	Spinal stenosis- Good	Spondylosis- Good	Scoliosis-Not Good Benign	syndromes)- Good	arthropathies- Not Good	areata-Not Good	joints-Not Good	nervous system-Good
			Gastritis and	Diaphragmatic		neoplasm of stomach and	Oesophagitis and		Barrett's	Infections of	
	Peptic ulcer	Peptic ulcer	duodenitis-	hernia-Not	Peritonitis-	duodenum-	oesophageal	Appendicitis-	oesophagus-	the digestive	Cholelithiasis-
7	disease	disease-Good	Good	Good	Good	Good	ulcer-Good	Good	Good	system-Good	Not Good
		Fibromatoses-	Retinal detachment-	Primary Malignancy Malignant Melanoma-	Benign essential tremor-Not	Disorders of autonomic nervous system-Not	Hyposplenism-	Primary Malignancy Other Skin and subcutaneous	Primary or Idiopathic Thrombocytop aenia-Not	Nasal polyp-	Subdural haematoma - nontraumatic-
8	Fibromatoses	Good	Good	Good	Good	Good	Good	tissue-Good	Good	Good	Not Good Delirium, not induced by alcohol and
									Primary		other
	For all 1	For all or	0-1	For the Claim	Carlinat At 1	Callana	Ellenene 1	Carpal tunnel	pulmonary	NA i - v	psychoactive
_	Fracture of	Fracture of wrist-Good	Osteoporosis- Good	Fracture of hip- Good		Collapsed vertebra-Good	Fibromatoses- Good	syndrome- Good	hypertension- Not Good	Menieres	substances- Good
5	wrist	wrist-dood	Good	Good	Good	vertebra-0000	Good	Good	NOT GOOD	disease-Good	Good

10	Ischaemic stroke	Ischaemic stroke-Good	Intracerebral haemorrhage- Good	Stroke Not otherwise specified (NOS)-Good	Transient ischaemic attack-Good	Aspiration pneumonitis-Good	Subarachnoid haemorrhage- Good	Subdural haematoma - nontraumatic- Good	Delirium, not induced by alcohol and other psychoactive substances-Good Primary	Bell's palsy-Not Good Primary	Syndrome of inappropriate secretion of antidiuretic hormone-Good
						Secondary	Secondary	Benign	Malignancy	Malignancy	Primary
	Primary	Primary	Primary	Postmenopaus	Primary	Malignancy	Malignancy	neoplasm and	Lung and	Malignant	Malignancy
	Malignancy	Malignancy	Malignancy	al bleeding-	Malignancy	Other organs-	Lymph Nodes-	polyp of uterus-	trachea-Not	Melanoma-	Thyroid-Not
11	Breast	Breast-Good	Uterine-Good	Good	Cervical-Good	Good	Good	Not Good	Good	Not Good	Good
			Pericardial	Pulmonary		Infections of	Secondary	Primary	Atrial	Primary	
	<b>5</b> 1 1	<b>5</b> 1 1	effusion	collapse (excl	Secondary	Other or	pulmonary	pulmonary	Fibrillation and	Malignancy	
4.2	Pleural	Pleural	(noninflammat	•	Malignancy	unspecified	hypertension-	hypertension-	flutter-Not	Other Organs-	Pleural plaque-
12	effusion	effusion-Good	ory)-Good	)-Good	Pleura-Good	organs-Good	Good	Good	Good	Good	Good
	Peripheral	Peripheral neuropathies						Postviral			
	neuropathies	(excluding						fatigue			
	(excluding	cranial nerve						syndrome,			
	cranial nerve	and carpal	Diabetic	Disorders of				neurasthenia			
	and carpal	tunnel	neurological	autonomic	Intervertebral	Carpal tunnel		and	Erectile	Peripheral	
	tunnel	syndromes)-	complications-	nervous	disc disorders-	syndrome-Not	Spinal stenosis-	fibromyalgia-	dysfunction-	arterial disease-	Spondylolisthe
13	syndromes)	Good	Good	system-Good	Good	Good	Good	Good	Good	Good	sis-Good
						Benign					
						neoplasm of					
						colon, rectum,				Postcoital and	
			Anorectal		Irritable bowel	anus and anal		Haemangioma	Alopecia	contact	
		Anal fissure-	fistula-Not	Pilonidal cyst-	syndrome-	canal-Not	Ulcerative	(any site)-Not	areata-Not	bleeding-Not	
14	Anal fissure	Good	Good	Not Good	Good	Good	colitis-Good	Good	Good	Good	Vitiligo-Good

15	Postmenopau sal bleeding Venous	Good Venous	polyp of uterus- Good	Primary Malignancy Uterine-Good	Endometrial hyperplasia and hypertrophy- Good	Leiomyoma of uterus-Good Primary	Benign neoplasm of ovary-Not Good	Primary Malignancy Breast-Good	Female genital prolapse-Good	0 ,	Postcoital and contact bleeding-Good
	thromboemb	thromboembol	Pulmonary	Bell's palsy-Not	Dlourel	Malignancy Other Organs	Fracture of him	Doritonitic	Chalangitis	Thrombophilia-	Atrial  Fibrillation and
16	olic disease (Excl PE)	ic disease (Excl PE)-Good	embolism- Good	Good	effusion-Good	Other Organs- Good	Fracture of hip- Good	Good	Cholangitis- Good	Good	flutter-Good
10	(LACIFL)	r L)-Good	Good	Good	emusion-good		Delirium, not induced by alcohol and	dood	Good	dood	nutter-good
						Subdural	other			S. 1.1	
	Fracture of	Fracture of hip-	Collansod	Ostoonorosis	Fracture of	haematoma -	psychoactive substances-	Dementia-	Hyporparathyr	Right bundle branch block-	Essential tremor-Not
17	hip	Good	•	Osteoporosis- Good	wrist-Good	nontraumatic- Not Good	Good	Good	Hyperparathyr oidism-Good	Not Good	Good
1,	· · · · ·	3004	Visual	Good	Wist Good	Retinal	3004	Diabetic	ciaisiii Good	Not Good	Diabetic
		Macular	impairment		Retinal	vascular		ophthalmic	Anterior and		neurological
	Macular	degeneration-	and blindness-		detachment-	occlusions-	Glaucoma-	complications-	Intermediate	Eye infections-	complications-
18	degeneration	Good	Good	Cataract-Good	Good	Good	Good	Good	Uveitis-Good	Not Good	Good
		Benign neoplasm of colon, rectum, anus and anal	Diverticular disease of intestine (acute and	Benign neoplasm of stomach and duodenum-	Angiodysplasia	Primary Malignancy colorectal and	Anorectal	Anal fissure-	Diaphragmatic hernia-Not	Barrett's oesophagus-	Peritonitis-
19	canal	canal-Good Postviral	chronic)-Good	Good	of colon-Good	anus-Good	fistula-Good Peripheral neuropathies	Good	Good	Good	Good
	Postviral	fatigue					(excluding				
	fatigue	syndrome,		Lupus			cranial nerve				
	syndrome,	neurasthenia		erythematosus			and carpal				
	neurasthenia	and		(local and		Irritable bowel	tunnel	Intervertebral		Neuromuscular	
20	and	fibromyalgia-	Sjogren's	systemic)-	Migraine-Not	syndrome-	syndromes)-	disc disorders-	Spondylolisthe	•	Leiomyoma of
20	fibromyalgia	Good	disease-Good	Good	Good	Good	Good	Good	sis-Good	bladder-Good	uterus-Good

Dysmenorrho 21 ea 22 Cholelithiasis	Dysmenorrhoe a-Good Cholelithiasis- Good	Endometriosis- Good Cholecystitis- Good	Leiomyoma of uterus-Good Cholangitis- Good	Menorrhagia and polymenorrho ea-Not Good Pancreatitis- Good	Postcoital and contact bleeding-Not Good Gastritis and duodenitis-Not Good	Carcinoma in situ cervical- Not Good Scleritis and episcleritis-Not Good	Endometrial hyperplasia and hypertrophy- Good  Appendicitis- Not Good	Polycystic ovarian syndrome- Good Diaphragmatic hernia-Not Good	Prematurity- Not Good Peptic ulcer disease-Not Good	Female pelvic inflammatory disease-Good  Obesity-Good
Primary Malignancy 23 Prostate	Primary Malignancy Prostate-Good	Secondary Malignancy Bone-Good	Primary Malignancy Bladder-Good	Hyperplasia of prostate-Good	Secondary Malignancy Other organs- Good	Obstructive and reflux uropathy-Not Good	Erectile dysfunction- Good	Asbestosis- Good	Secondary malignancy Liver and intrahepatic bile duct-Good	Hyposplenism- Good
Myocardial 24 Infarction	Myocardial Infarction- Good	Unstable Angina-Good	Stable Angina- Good	Coronary heart disease not otherwise specified-Good	Left bundle branch block-	Heart failure- Good	Atrioventricula r block, second degree-Good	Ventricular tachycardia- Good	Atrioventricula r block, first degree-Good	Atrioventricula r block, complete- Good
Peripheral	Peripheral	Abdominal Aortic	Diabetic neurological		Retinal vascular	Transient		Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Diabetes Mellitus,Äì	Visual impairment	Coronary heart disease not
arterial 25 disease	arterial disease Good	- Aneurysm- Good	complications- Good Benign neoplasm of	Unstable Angina-Good Gastro-	occlusions- Good	ischaemic attack-Good	Stable Angina- Good  Oesophagitis	other or not specified-Good Diverticular disease of intestine	and blindness- Good Benign neoplasm of colon, rectum,	otherwise specified-Good
Diaphragmati 26 chernia	Diaphragmatic hernia-Good	Gastritis and duodenitis- Good	stomach and duodenum-Not Good	oesophageal reflux disease- Good	Barrett's oesophagus- Good	Peptic ulcer disease-Good	and oesophageal ulcer-Good	(acute and chronic)-Not Good	anus and anal canal-Not Good	Cholelithiasis- Not Good

Tubulo-Obstructive interstitial Primary Urolithiasis-Hyperplasia of Appendicitis-Bell's palsy-Not Cholecystitisand reflux nephritis-Not Non-acute Malignancy Pancreatitis-27 Urolithiasis prostate-Good Not Good Good Good Good uropathy-Good Good cystitis-Good Bladder-Good Not Good Diabetic Retinal Visual vascular Macular Anterior and ophthalmic End stage impairment Glaucomaocclusionsand blindnessdegeneration-Intermediate complications-Parkinson's renal disease-Good Good Good **Uveitis-Good** Good 28 Glaucoma Good Cataract-Good Keratitis-Good Good disease-Good Ear and Upper Infection of Infection of Lower Respiratory Respiratory other or skin and **Urinary Tract** Tract Infections of Female pelvic Tract unspecified Infection of subcutaneous Infection of Infections-Not the digestive Urinary Tract Infections-Infections-**Acute Kidney** inflammatory genitourinary male genital tissues-Not bones and 29 Infections Good Good system-Good disease-Good Good system-Good Injury-Good ioints-Good system-Good Good Peripheral neuropathies (excluding Primary cranial nerve Malignancy and carpal Raynaud's Haemangioma Seborrheic Malignant Irritable bowel tunnel syndrome-Not Chronic Hearing loss-Menieres syndrome-(any site)dermatitis-Not Melanomasyndromes)-Tinnitus-Good Good 30 Tinnitus Good Good Good Good Good sinusitis-Good Not Good disease-Good Infection of Lower other or Ear and Upper Lower Respiratory unspecified Infections of Infection of Infection of Respiratory Infections of Urinary Tract Respiratory Tract genitourinary Tract the digestive bones and male genital Tract Infectionsthe Heart-Not Infection of system-Not Infections-Not Infectionssystem-Not joints-Not system-Not Agranulocytosi 31 Infections liver-Good Good Good Good Good Good Good s-Good Good Good Delirium, not induced by alcohol and Stroke Not Retinal other Subarachnoid Transient Transient Intracerebral vascular Peripheral Nonrheumatic otherwise psychoactive specified occlusions-Bell's palsy-Not arterial disease- substancesischaemic ischaemic Ischaemic haemorrhagehaemorrhageaortic valve 32 attack attack-Good stroke-Good (NOS)-Good Good Good Not Good Good Good Good disorders-Good

Bacterial Diseases (exc 33 TB)	Bacterial Diseases (excl TB)-Good	Other or unspecified infectious organisms- Good	Viral diseases (excl chronic hepatitis/HIV)- Not Good Monoclonal gammopathy of	Parasitic infections-Not Good	Tubulo- interstitial nephritis-Good	Female infertility-Good	Obstructive and reflux uropathy-Good	Acute Kidney Injury-Good	Aspiration pneumonitis-Good	Infections of Other or unspecified organs-Good
Polymyalgia 34 Rheumatica Primary	Polymyalgia Rheumatica- Good	Giant Cell arteritis-Good	undetermined significance (MGUS)-Not Good	Rheumatoid Arthritis-Not Good	Collapsed vertebra-Good	Spinal stenosis- Good	Epilepsy-Good	Spondylolisthe sis-Good	Menieres disease-Good	Nonrheumatic aortic valve disorders-Good
Malignancy Other Skin and	Primary Malignancy Other Skin and subcutaneous tissue-Good	Actinic keratosis-Good	Primary Malignancy Malignant Melanoma-	Haemangioma (any site)-Not Good	Primary Malignancy Oropharyngeal- Not Good	- Fibromatoses- Good	Rosacea-Not Good	Seborrheic dermatitis-Not Good	Hyperplasia of prostate-Not	Gout-Good
36 Urticaria	Urticaria-Good	Dermatitis (atopc/contact /other/unspeci fied)-Good	Seborrheic	Keratitis-Good	Chronic sinusitis-Good	Lichen planus- Not Good	Hidradenitis suppurativa- Not Good	Benign neoplasm of ovary-Not Good	Alopecia areata-Good	Raynaud's syndrome-Not Good
Actinic 37 keratosis	Actinic keratosis-Good	Primary Malignancy Other Skin and subcutaneous tissue-Good	Primary Malignancy Malignant Melanoma- Good	Haemangioma (any site)-Not Good	Hearing loss- Not Good	Osteoarthritis (excl spine)- Good	Fibromatoses- Not Good	Seborrheic dermatitis- Good	Rosacea-Good	Enthesopathie s & synovial disorders-Not Good
38 Rosacea	Rosacea-Good	Acne-Not Good	Seborrheic dermatitis- Good	Primary Malignancy Other Skin and subcutaneous tissue-Good	Ulcerative colitis-Good	Migraine-Good	Actinic keratosis-Good	Vitiligo-Not Good	Urticaria-Not Good	Female genital prolapse-Not Good

39	Osteoporosis			Fracture of wrist-Good	Fracture of hip- Good Gastro-	Scoliosis-Good Benign neoplasm of	Primary Malignancy Breast-Good Oesophagitis	Spondylosis- Good	Cataract-Good	Osteoarthritis (excl spine)- Not Good	Hyperparathyr oidism-Good  Infections of
	Gastritis and	Gastritis and duodenitis-	Diaphragmatic hernia-Not	Peptic ulcer	oesophageal reflux disease-	stomach and duodenum-	and oesophageal	Irritable bowel syndrome-	Cholelithiasis-	Iron deficiency	Other or
40	duodenitis	Good	Good Benign	disease-Good	Good	Good	ulcer-Good	Good	Not Good	anaemia-Good	•
	Diverticular disease of	Diverticular disease of	neoplasm of colon, rectum,			Abdominal			Benign neoplasm of		
	intestine	intestine	anus and anal canal-Not	Peritonitis-		Anguniam	Diaphragmatic		stomach and duodenum-Not	Osteoarthritis	Gastritis and duodenitis-Not
41	(acute and chronic)	(acute and chronic)-Good	Good	Good	syndrome-Not Good	Good	hernia-Not Good	prolapse-Not Good	Good	Not Good	Good
	,	,			Peripheral neuropathies (excluding cranial nerve						
					and carpal	Menorrhagia					
		Carpal tunnel	Enthesopathie	6 11 .	tunnel	and	Intervertebral	Osteoarthritis			E1 .
12	Carpal tunnel syndrome	syndrome- Good	s & synovial disorders-Good	Spondylosis-	syndromes)- Not Good	polymenorrho ea-Not Good	disc disorders- Not Good	(excl spine)- Good	Fracture of wrist-Good	Obesity-Good	Fibromatoses- Not Good
72	syndrome	doou	uisorders dood	Not dood	Not dood	ea-Not Good	Not Good	dodd	Wist Good	Primary	Not dood
		Seborrheic	Dermatitis (atopc/contact					Chronic		Malignancy Other Skin and subcutaneous	Enthesopathie s & synovial
	Seborrheic	dermatitis-	/other/unspeci		Psoriasis-Not	Urticaria-Not		sinusitis-Not	Tinnitus-Not	tissue-Not	disorders-Not
43	dermatitis	Good	fied)-Not Good	Good	Good	Good	Acne-Not Good	Good	Good	Good Transient	Good
			Intervertebral	Osteoarthritis		Carpal tunnel		Enthesopathie		ischaemic	Diaphragmatic
		Spondylosis-	disc disorders-	(excl spine)-	Spinal stenosis-	•	Spondylolisthe	•	Osteoporosis-	attack-Not	hernia-Not
44	Spondylosis	Good	Good	Good	Good	Good	sis-Good	disorders-Good	Good	Good	Good

Other or unspecified infectious 45 organisms	Other or unspecified infectious organisms- Good	Bacterial Diseases (excl TB)-Good	Viral diseases (excl chronic hepatitis/HIV)- Good	Infections of Other or unspecified organs-Good	Aspiration pneumonitis- Good	Pleural effusion-Good	Mycoses-Good	Parasitic infections- Good	Pulmonary collapse (excl pneumothorax )-Good	Acute Kidney Injury-Good
Vitamin B12	Vitamin B12	Folate	Iron deficiency	Other		Myelodysplast				
deficiency	deficiency	deficiency	anaemia-Not	anaemias-Not	Coeliac disease-	•	Dementia-		Rheumatoid	
46 anaemia	anaemia-Good	anaemia-Good	Good	Good	Good	Good	Good	Vitiligo-Good	Arthritis-Good	Psoriasis-Good
		Primary	Primary						Primary Malignancy Other Skin and	Diverticular disease of intestine
		Malignancy	Malignancy	Erectile		Hydrocoele	Right bundle	Abdominal	subcutaneous	(acute and
Hyperplasia	Hyperplasia of	Prostate-Not	Bladder-Not	dysfunction-	Urolithiasis-	(incl infected)-	branch block-	Hernia-Not	tissue-Not	chronic)-Not
47 of prostate	prostate-Good	Good	Good	Not Good	Good	Not Good	Good	Good Benign	Good	Good
			Folate	Vitamin B12	Menorrhagia			neoplasm of		
Iron		Other	deficiency	deficiency	and	Gastritis and		stomach and		
deficiency	Iron deficiency		anaemia-Not	anaemia-Not	polymenorrho	duodenitis-	Leiomyoma of	duodenum-	Peptic ulcer	Anxiety
48 anaemia	anaemia-Good	Good	Good	Good	ea-Good	Good	uterus-Good	Good	disease-Good	disorders-Good
			_	Benign						Primary
Female			Postmenopaus	•	Urinary	Neuromuscular		Sjogren's		Malignancy
genital	Female genital		al bleeding-	polyp of uterus-		dysfunction of	contact	disease-Not	Leiomyoma of	Breast-Not
49 prolapse	prolapse-Good	prolapse-Good	Good	Good	Good	bladder-Good	bleeding-Good	G000	uterus-Good Enthesopathie	Good
	Urinary	Neuromuscular				Osteoarthritis	Allergic and		s & synovial	
Urinary	Incontinence-	dysfunction of	Female genital	Hearing loss-	Depression-	(excl spine)-	=	Dementia-	disorders-Not	
50 Incontinence	Good	bladder-Good	prolapse-Good	=	Good	Good	Good	Good	Good	Obesity-Good
			, ,			Oesophagitis	Dermatitis			Primary
	Other		Folate		Vitamin B12	and	(atopc/contact			Malignancy
Other	anaemias-	Iron deficiency	deficiency	Depression-	deficiency	oesophageal	/other/unspeci	Peptic ulcer	Heart failure-	colorectal and
51 anaemias										

Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Diabetic Diabetic Diabetes Retinal Visual Diabetic ophthalmic Mellitus ,Äì Macular Erectile vascular Retinal neurological impairment ophthalmic complicationscomplicationsother or not degenerationdysfunctionocclusionsand blindness- Glaucomadetachment-52 complications Good Good specified-Good Good Good Good Good Obesity-Good Good Good Primary Coronary heart Atrioventricula Nonrheumatic Left bundle Other Myocardial Nonrheumatic Dilated disease not r block. pulmonary Heart failure-Cardiomyopat Infarctionmitral valve branch blockaortic valve cardiomyopath otherwise completehypertension-53 Heart failure Good hy-Good disorders-Good Good Good disorders-Good y-Good specified-Good Good Not Good Delirium, not induced by alcohol and Stroke Not Stroke Not other Subarachnoid otherwise otherwise Transient Intracerebral Aspiration psychoactive haemorrhage- Heart failurespecified specified Ischaemic ischaemic pneumonitissubstanceshaemorrhage-Dementia-54 (NOS) (NOS)-Good stroke-Good attack-Good Good Good Good Epilepsy-Good Good Good Good Diverticular disease of intestine Gastro-Irritable Irritable bowel (acute and Allergic and Gastritis and oesophageal bowel syndrome-Chronic chronic)-Not Anxiety chronic rhinitis- Anal fissureduodenitis-Not Dysmenorrhoe reflux disease-Not Good 55 syndrome sinusitis-Good Good disorders-Good Good Not Good Migraine-Good Good a-Good Good

56	Gout Oesophagitis	Gout-Good Oesophagitis	Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Diabetes Mellitus ,Äì other or not specified-Good	Primary Malignancy Other Skin and subcutaneous tissue-Good	Osteoarthritis (excl spine)- Good Gastro-	Heart failure- Good	Glomerulonep hritis-Good	Erectile dysfunction- Good Benign neoplasm of	Actinic keratosis-Not Good	Enthesopathie s & synovial disorders-Not Good	Fibromatoses- Not Good Lower Respiratory
	and	and	Diaphragmatic	Barrett's	oesophageal	Gastritis and		stomach and	Other		Tract
	oesophageal	oesophageal	hernia-Not	oesophagus-	reflux disease-	duodenitis-	Peptic ulcer	duodenum-	anaemias-	Depression-	Infections-
57	ulcer	ulcer-Good	Good	Good	Good	Good	disease-Good	Good	Good	Good	Good
	Erectile dysfunction	Erectile dysfunction- Good Dementia-	Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Diabetes Mellitus ,Äì other or not specified-Good Delirium, not induced by alcohol and other psychoactive substances-	Parkinson's	Primary Malignancy Prostate-Good  Aspiration pneumonitis-	Diabetic ophthalmic complications- Good	Dyslipidaemia- Good Left bundle branch block-	Peripheral neuropathies (excluding cranial nerve and carpal tunnel syndromes)- Good  End stage renal disease-	Sleep apnoea- Good  Stroke Not otherwise specified	Fibromatoses- Good	Abdominal Hernia-Not Good Intracerebral haemorrhage-
59	Dementia	Good	Not Good	disease-Good	Good	Good	Good	Good	(NOS)-Good	stroke-Good	Good

	Alcohol	Alcohol Problems-	Liver fibrosis, sclerosis and	Hepatic failure-	Alcoholic liver	Fatty Liver-	Other psychoactive substance	Portal hypertension-	Pancreatitis-	Oesophageal	Personality
60	Problems	Good	cirrhosis-Good	Good	disease-Good Schizophrenia,	Good	misuse-Good	Good	Good	varices-Good	disorders-Good
	Other	Other			schizotypal		Anorexia and	Bipolar			
	psychoactive	psychoactive			and delusional	Obsessive-	bulimia	affective	Alcohol	Alopecia	Crohn's
	substance	substance	Chronic viral	Personality	disorders-Not	compulsive	nervosa-Not	disorder and	Problems-	areata-Not	disease-Not
61	misuse	misuse-Good	hepatitis-Good	disorders-Good	Good	disorder-Good	Good	mania-Good	Good	Good	Good
					Postcoital and		Polycystic			Menorrhagia	
					contact	Carcinoma in	ovarian	Hidradenitis		and	
			Rosacea-Not	Dysmenorrhoe	bleeding-Not	situ cervical-	syndrome-	suppurativa-	Prematurity-	polymenorrho	Crohn's
62	Acne	Acne-Good	Good	a-Good	Good	Not Good	Good	Not Good	Not Good	ea-Not Good	disease-Good Bipolar
						Dermatitis				Gastro-	affective
			Psoriatic	Seborrheic		(atopc/contact		Hidradenitis		oesophageal	disorder and
			arthropathy-	dermatitis-Not	Rheumatoid	/other/unspeci		suppurativa-	Urticaria-Not	reflux disease-	mania-Not
63	Psoriasis	Psoriasis-Good	Good	Good	Arthritis-Good	fied)-Good	Obesity-Good	Good	Good	Not Good	Good
				Hypertrophy of							
			Allergic and	nasal	Irritable bowel				Seborrheic		
	Chronic	Chronic	chronic rhinitis-	turbinates-	syndrome-	Nasal polyp-		Migraine-Not	dermatitis-Not	Dysmenorrhoe	
64	sinusitis	sinusitis-Good	Good	Good	Good	Good	Asthma-Good	Good	Good	a-Not Good	Urticaria-Good
							Primary				
							Malignancy				
							Other Skin and			Enthesopathie	
			Hydrocoele	Erectile			subcutaneous		Actinic	s & synovial	
	Abdominal	Abdominal	(incl infected)-	dysfunction-	Hyperplasia of	Appendicitis-	tissue-Not	Hearing loss-	keratosis-Not	disorders-Not	Psoriasis-Not
65	Hernia	Hernia-Good	Good	Not Good	prostate-Good	Not Good	Good	Not Good	Good	Good	Good

			Type 1								
			Diabetes								
			Mellitus, Type								
			2 Diabetes								
			Mellitus, and								
			Diabetes		Polycystic			Menorrhagia	Diabetic		
			Mellitus ,Äì		ovarian		Hidradenitis	and	ophthalmic	Allergic and	Osteoarthritis
			other or not	Sleep apnoea-	syndrome-	Dyslipidaemia-	suppurativa-	polymenorrho	complications-	chronic rhinitis-	(excl spine)-
66 Ol	besity	Obesity-Good	specified-Good	Good	Good	Good	Good	ea-Good	Good	Good	Good
				Enthesopathie			Dermatitis		Visual		
				s & synovial	Osteoarthritis	Urinary	(atopc/contact	•	impairment	Actinic	Seborrheic
		Hearing loss-		disorders-Not	(excl spine)-	Incontinence-	•		and blindness-	keratosis-Not	dermatitis-
67 He	earing loss	Good	Tinnitus-Good	Good	Good	Good	fied)-Not Good		Not Good	Good	Good
								Postviral			
								fatigue			
								syndrome,			
			Menorrhagia					neurasthenia		Enthesopathie	
			and		Irritable bowel	Chronic		and		s & synovial	Allergic and
			polymenorrho	Dysmenorrhoe	syndrome-	sinusitis-Not		fibromyalgia-		disorders-Not	chronic rhinitis-
68 M	ligraine	Migraine-Good	ea-Good	a-Good	Good	Good	Acne-Good	Not Good	Asthma-Good	Good	Not Good
						Endometrial					
M	Ienorrhagia	Menorrhagia				hyperplasia		Benign			
an	nd	and			Postcoital and	and		neoplasm and	Carcinoma in		Benign
ро	olymenorrh	polymenorrho	Leiomyoma of	Dysmenorrhoe	contact	hypertrophy-	Endometriosis-	polyp of uterus-	situ cervical-		neoplasm of
69 oe	ea	ea-Good	uterus-Good	a-Good	bleeding-Good	Good	Good	Good	Good	Migraine-Good	ovary-Good

	Type 1										
	Diabetes	Type 1									
	Mellitus,	Diabetes									
	Type 2	Mellitus, Type									
	Diabetes	2 Diabetes									
	Mellitus, and	Mellitus, and									
	Diabetes	Diabetes	Diabetic				Diabetic				
	Mellitus ,Äì	Mellitus ,Äì	ophthalmic		Erectile		neurological	Peripheral			Osteoarthritis
	other or not	other or not	complications-	Dyslipidaemia-	dysfunction-		complications-	arterial disease-		Hypertension-	(excl spine)-
70	specified	specified-Good	Good	Good	Good	Obesity-Good	Good	Good	Gout-Good	Good	Good
					Visual	5 · · · · ·		Diabetic	Atrial		
			Macular		impairment	Retinal	Anterior and	ophthalmic	Fibrillation and		
74	<b>.</b>		degeneration-	Glaucoma-	and blindness-	detachment-	Intermediate	complications-	flutter-Not	Osteoporosis-	Hearing loss-
/1	Cataract	Cataract-Good	Good	Good	Good	Good	Uveitis-Good	Good	Good	Good	Good
					Coronary heart						
				Myocardial	disease not			Peripheral	Nonrheumatic	Transient	Left bundle
		Stable Angina-	Unstable	Infarction-	otherwise	Dyslipidaemia-	Heart failure-	arterial disease-	aortic valve	ischaemic	branch block-
72	Stable Angina	•	Angina-Good	Good	specified-Good	• •	Good	Good	disorders-Good	attack-Good	Good
	S		J		•				Benign		
		Gastro-			Oesophagitis				neoplasm of		
	Gastro-	oesophageal		Gastritis and	and		Barrett's	Irritable bowel	stomach and	Allergic and	Chronic
	oesophageal	reflux disease-	Diaphragmatic	duodenitis-	oesophageal	Anxiety	oesophagus-	syndrome-	duodenum-	chronic rhinitis-	sinusitis-Not
73	reflux disease	Good	hernia-Good	Good	ulcer-Good	disorders-Good	Good	Good	Good	Good	Good
	Coronary	Coronary heart									
	heart disease	disease not	Myocardial				Atrial	Left bundle		Peripheral	Atrioventricula
			•							•	

74 specified

specified-Good Good

Good

Angina-Good

Good

flutter-Good

Good

Good

Good

degree-Good

75	Chronic obstructive pulmonary disease (COPD)	Chronic obstructive pulmonary disease (COPD)-Good	Asthma-Not Good	Respiratory failure-Good	Lower Respiratory Tract Infections-Not Good	Bronchiectasis- Not Good	Pneumothorax- Good	Primary Malignancy Lung and trachea-Good	Heart failure- Good	Other interstitial pulmonary diseases with fibrosis-Not Good	Abdominal Aortic Aneurysm- Good
			Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and								
			Diabetes	Enthesopathie	0-4		Coronary heart			For all a	Allanaia and
	Dyslipidaemi	Dyslipidaemia-	Mellitus – other or not	s & synovial disorders-Not	Osteoarthritis (excl spine)-	Stable Angina-	disease not otherwise		Hypertension-	Erectile dysfunction-	Allergic and chronic rhinitis-
76	a	Good	specified-Good		Good	Good	specified-Good	Obesity-Good	Good	Good	Good
			.,		Primary		.,	, , , , , , , , , , , , , , , , , , , ,			
					Malignancy		Stroke Not				
				Autism and	Brain, Other		otherwise		Other		
				Asperger's	CNS and	Intracerebral	specified	Aspiration	psychoactive	Subarachnoid	Alcohol
			Intellectual	syndrome-	Intracranial-	haemorrhage-	(NOS)-Not	pneumonitis-	substance	haemorrhage-	Problems-
77	Epilepsy	Epilepsy-Good	disability-Good	Good	Good	Good	Good	Good	misuse-Good	Good	Good
							Dermatitis		Benign neoplasm of brain and other		
	Hypo or	Hypo or		Primary	Atrial		(atopc/contact	Osteoarthritis	parts of central		
		hyperthyroidis	Hypertension-	Malignancy	Fibrillation and	Dyslipidaemia-	•		nervous		
78	sm	m-Good	Good	Thyroid-Good Dermatitis	flutter-Good	Good Enthesopathie	fied)-Good	Not Good	system-Good	Obesity-Good Hypertrophy of	Asthma-Good
	Allergic and	Allergic and		(atopc/contact		s & synovial	Irritable bowel		Osteoarthritis	nasal	
	chronic	chronic rhinitis-	Chronic	/other/unspeci		disorders-Not	syndrome-	Nasal polyp-	(excl spine)-	turbinates-	Hearing loss-
79	rhinitis	Good	sinusitis-Good	fied)-Good	Asthma-Good	Good	Good	Good	Not Good	Good	Good

80	Atrial Fibrillation and flutter	Atrial Fibrillation and flutter-Good	Pleural effusion-Not Good	Coronary heart disease not otherwise specified-Good	•	Heart failure- Good	Hypertension- Not Good	Primary pulmonary hypertension- Good	Cataract-Good	Nonrheumatic mitral valve disorders-Good Diverticular disease of	Nonrheumatic aortic valve disorders-Not Good
81	Osteoarthritis (excl spine)	Osteoarthritis (excl spine)- Good	Enthesopathie s & synovial disorders-Good Chronic	Spondylosis- Good	Dermatitis (atopc/contact /other/unspeci fied)-Not Good	•	Dyslipidaemia- Good	Allergic and chronic rhinitis- Not Good	Hypertension- Good	intestine (acute and chronic)-Not Good	Gastro- oesophageal reflux disease- Good
82	Asthma	Asthma-Good	obstructive pulmonary disease (COPD)-Not Good	Allergic and chronic rhinitis- Good Gastro- oesophageal	sinusitis-Good Irritable bowel	Dermatitis (atopc/contact /other/unspeci fied)-Good Dermatitis (atopc/contact	Good	Migraine-Good Allergic and		Obesity-Good Other psychoactive	Menorrhagia and polymenorrho ea-Not Good Enthesopathie s & synovial
83	Anxiety disorders	Anxiety disorders-Good	Depression- Good	reflux disease- Good	syndrome- Good	/other/unspeci fied)-Good	chronic sinusitis-Good	chronic rhinitis- Good	Iron deficiency anaemia-Good	substance misuse-Good	disorders-Not Good
84	Enthesopathi es & synovial disorders Dermatitis (atopc/contac	Enthesopathie s & synovial disorders-Good Dermatitis (atopc/contact	(excl spine)- Good	Dermatitis (atopc/contact /other/unspeci fied)-Not Good Enthesopathie s & synovial	chronic rhinitis-	Carpal tunnel syndrome-Not Good Osteoarthritis	Hearing loss- Not Good	Dyslipidaemia- Good	Depression- Not Good	Spondylosis- Good	Migraine-Not Good
85	t/other/unsp ecified)	/other/unspeci fied)-Good	Good	disorders-Not Good Dermatitis (atopc/contact		(excl spine)- Not Good Enthesopathie s & synovial	Urticaria-Good	Depression- Good Irritable bowel	Hypertension- Good Allergic and	Asthma-Good Osteoarthritis	Hearing loss- Good Urinary
86	Depression	Depression- Good	Anxiety disorders-Good	/other/unspeci fied)-Good	anaemias- Good	disorders-Not Good	Chronic sinusitis-Good	syndrome- Good	chronic rhinitis- Good	(excl spine)- Good	Incontinence- Good

Type 1 Diabetes Mellitus, Type 2 Diabetes

Mellitus, and

Enthesopathie Dermatitis Diabetes (atopc/contact Osteoarthritis Mellitus ,Äì s & synovial Atrial Hypo or Hypertension-/other/unspeci (excl spine)-Fibrillation and Dyslipidaemia- other or not disorders-Not Depressionhyperthyroidis Acute Kidney 87 Hypertension Good fied)-Good Good flutter-Good Good specified-Good Good Good m-Good Injury-Good

# **Hyperparameter Tuning for RETAIN and Deepr**

In this section, we performed Bayesian Optimisation to search for optimal parameters for Deepr (Table S4) and RETAIN (Table S5).

**Table S4.** Deepr Best Model for Subsequent Visit Prediction Task

Iteration	Filters	Kernel Size	FC I	FC II	FC III	Dropout I	Dropout II	Dropout III	Learning Rate	Average Precision
1	37	7	10	46	28	0.4139	0.4997	0.3718	0.0004	0.2599
2	24	5	28	19	13	0.4936	0.1722	0.4643	0.0062	0.2319
3	17	7	16	48	40	0.2250	0.3945	0.3264	0.0310	0.1815
4	33	7	6	10	25	0.1602	0.4476	0.3544	0.0026	0.2640
5	32	7	4	30	32	0.3714	0.3382	0.3573	0.0353	0.2005
6	13	4	50	17	47	0.3424	0.1830	0.3056	0.0008	0.3274
7	12	3	3	6	43	0.4201	0.2387	0.3884	0.0123	0.2112
8	30	7	16	26	41	0.2055	0.3343	0.1541	0.0011	0.3256
9	35	5	50	24	25	0.1567	0.4056	0.1329	0.0004	0.3433
10	41	7	9	35	19	0.4885	0.3548	0.3080	0.0004	0.2343
11	4	4	33	12	39	0.1811	0.1251	0.1066	0.0010	0.3051
12	48	4	36	45	37	0.2143	0.3486	0.1222	0.0015	0.3504
13	36	3	39	10	48	0.1992	0.4164	0.1183	0.0050	0.3291
14	45	4	44	40	26	0.2329	0.2900	0.1280	0.0903	0.0042
15	40	4	35	48	11	0.1600	0.4704	0.1211	0.0019	0.3125
16	49	3	47	41	40	0.1002	0.2541	0.1284	0.0019	0.3588
17	47	3	50	47	12	0.2519	0.2125	0.1668	0.0005	0.2988
18	47	3	37	35	50	0.1059	0.4225	0.1120	0.0034	0.3487
19	47	4	48	34	39	0.2177	0.3607	0.1301	0.0016	0.3567
20	46	3	46	45	43	0.2007	0.1609	0.1196	0.0044	0.3356



**Table S5.** RETAIN Best Model for Subsequent Visit Prediction Task

Iteration	Embedding Size	Recurrent Size	Dropout Embedding	Dropout Context	L2	Average Precision
1	142	90	0.3846	0.0224	0.0891	0.1822
2	124	43	0.3922	0.0382	0.0003	0.1815
3	173	90	0.3929	0.2238	0.0014	0.3479
4	145	91	0.4117	0.0404	0.0102	0.2049
5	153	92	0.4569	0.0642	0.0116	0.2049
6	120	37	0.4335	0.4017	0.0728	0.1815
7	180	102	0.3567	0.3307	0.0039	0.2469
8	195	38	0.3805	0.2711	0.0010	0.3740
9	165	119	0.4969	0.1964	0.0047	0.2292
10	174	92	0.3862	0.2078	0.0019	0.3329
11	145	110	0.3928	0.1926	0.0891	0.1813
12	195	83	0.4418	0.4787	0.0011	0.3543
13	187	110	0.3456	0.2083	0.0123	0.2122
14	144	80	0.3717	0.0932	0.0032	0.3038
15	193	68	0.4528	0.3950	0.0022	0.3280
16	198	45	0.4344	0.3324	0.0442	0.1828
17	145	64	0.4213	0.1950	0.0626	0.1813
18	171	116	0.4166	0.4950	0.0028	0.3197
19	186	38	0.4062	0.4916	0.0011	0.3309
20	136	54	0.3503	0.1678	0.0067	0.2162

#### **Disease-wise Model Performance**

In this section, we show the disease-wise BEHRT performance in terms of AUROC and APS for the next 6 months prediction task. We have displayed codes with occurrence ratio 0.01 or higher. And detailed below is the description of the Caliber code and the chapter along with the APS/AUROC.

**Table S6.** Diseases Prediction Performance

Caliber	APS	AUROC	Description	Ratio	Caliber Chapter
92	0.066765	0.723828	Gastritis and duodenitis	0.011198	Diseases of the digestive system
66	0.108185	0.797633	Diaphragmatic hernia	0.011490	Diseases of the digestive system
100	0.118093	0.742646	Hearing loss	0.021964	Diseases of the ear and mastoid process
273	0.132567	0.773249	Spondylosis	0.013459	Diseases of the musculoskeletal system and connective tissue
189	0.142594	0.844596	Pleural effusion	0.010229	Diseases of the respiratory system
172	0.162186	0.798654	Other anaemias	0.023303	Diseases of the blood and blood- forming organs and certain disor- ders involving the immune mech- anism
29	0.163355	0.822707	Bacterial Diseases (excl TB)	0.023979	Certain infectious and parasitic diseases
130	0.167457	0.804545	Iron deficiency anaemia	0.020780	Diseases of the blood and blood- forming organs and certain disor- ders involving the immune mech- anism
295	0.169674	0.836610	Urinary Tract Infections	0.022534	Diseases of the genitourinary system



**Table S6.** Diseases Prediction Performance

69	0.170852	0.811759	Diverticular disease of intestine	0.015966	Diseases of the digestive system
8	0.177170	0.810068	(acute and chronic) Allergic and chronic rhinitis	0.023241	Diseases of the respiratory sys-
170	0.181182	0.847938	Osteoporosis	0.013982	tem Diseases of the musculoskeletal system and connective tissue
71	0.184000	0.790655	Dyslipidaemia	0.026010	Endocrine, nutritional and metabolic diseases
168	0.184629	0.799592	Oesophagitis and oesophageal ulcer	0.022426	Diseases of the digestive system
217	0.203218	0.855593	Primary Malignancy Other Skin and subcutaneous tissue	0.012013	Neoplasms
93	0.206455	0.776367	Gastro-oesophageal reflux disease	0.026271	Diseases of the digestive system
290	0.208524	0.814481	Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus, and Diabetes Mellitus – other or not specified	0.021226	Endocrine, nutritional and metabolic diseases
3	0.219637	0.869366	Actinic keratosis	0.012490	Diseases of the skin and subcutaneous tissue
131	0.220017	0.874319	Irritable bowel syndrome	0.011182	Diseases of the digestive system
294	0.223863	0.824114	Urinary Incontinence	0.020057	Diseases of the genitourinary system
169	0.234444	0.785766	Osteoarthritis (excl spine)	0.043714	Diseases of the musculoskeletal system and connective tissue
176	0.245906	0.834062	Other or unspecified infectious organisms	0.030963	Diseases of the respiratory system
95	0.249208	0.894444	Glaucoma	0.011367	Diseases of the eye and adnexa
109	0.250573	0.885547	Hyperplasia of prostate	0.020842	Diseases of the genitourinary system
184	0.264687	0.879325	Peripheral arterial disease	0.010951	Diseases of the circulatory system
79	0.265863	0.762939	Enthesopathies & synovial disorders	0.047036	Diseases of the musculoskeletal system and connective tissue
81	0.267187	0.905812	Erectile dysfunction	0.017873	Mental and behavioural disorders
140	0.268504	0.867094	Lower Respiratory Tract Infections	0.023518	Certain infectious and parasitic diseases
63	0.271027	0.753816	Dermatitis (atopc/contact/other/unspecified)	0.049051	Diseases of the skin and subcutaneous tissue
142	0.292598	0.893802	Macular degeneration	0.010752	Diseases of the eye and adnexa
274	0.296798	0.889236	Stable Angina	0.032039	Diseases of the circulatory system
57	0.301041	0.900088	Coronary heart disease not otherwise specified	0.035177	Diseases of the circulatory system
275	0.307238	0.911618	Stroke Not otherwise specified (NOS)	0.023395	Diseases of the nervous system
45	0.319433	0.863447	Cataract	0.042099	Diseases of the eye and adnexa
1	0.319972	0.845171	Abdominal Hernia	0.019180	Diseases of the digestive system
44	0.325143	0.843480	Carpal tunnel syndrome	0.012013	Diseases of the nervous system
101	0.334902	0.912117	Heart failure	0.024918	Diseases of the circulatory system
164	0.335131	0.879670	Obesity	0.017442	Endocrine, nutritional and metabolic diseases
'	'	'			

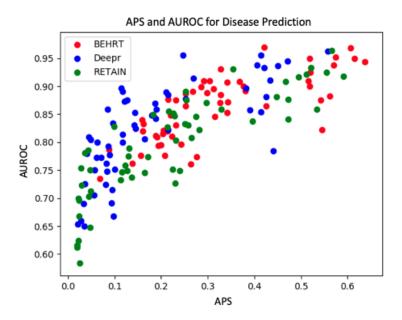
**Table S6.** Diseases Prediction Performance

22	0.348523	0.885055	Asthma	0.026133	Diseases of the respiratory sys-
					tem
97	0.349361	0.882694	Gout	0.018058	Diseases of the musculoskeletal system and connective tissue
65	0.350132	0.942604	Diabetic ophthalmic complica-	0.018919	Endocrine, nutritional and metabolic diseases
147	0.368465	0.911541	Migraine	0.012028	Diseases of the nervous system
229	0.398842	0.904686	Psoriasis	0.011751	Diseases of the musculoskeletal system and connective tissue
17	0.410899	0.858498	Anxiety disorders	0.041914	Mental and behavioural disorders
146	0.433645	0.969406	Menorrhagia and polymenor- rhoea	0.015504	Diseases of the genitourinary system
113	0.489456	0.905032	Hypo or hyperthyroidism	0.047897	Endocrine, nutritional and metabolic diseases
302	0.491672	0.855823	Vitamin B12 deficiency anaemia	0.014489	Diseases of the blood and blood- forming organs and certain disor- ders involving the immune mech- anism
51	0.501496	0.923082	Chronic obstructive pulmonary disease (COPD)	0.036869	Diseases of the respiratory system
23	0.514881	0.901268	Atrial Fibrillation and flutter	0.077629	Diseases of the circulatory system
110	0.531597	0.819527	Hypertension	0.200618	Diseases of the circulatory system
61	0.542223	0.950442	Dementia	0.024656	Mental and behavioural disorders
62	0.553561	0.877904	Depression	0.076876	Mental and behavioural disorders
85	0.573880	0.934049	Female genital prolapse	0.015781	Diseases of the genitourinary system
220	0.575574	0.964776	Primary Malignancy Prostate	0.011844	Neoplasms
6	0.583305	0.952656	Alcohol Problems	0.014535	Mental and behavioural disorders
194	0.647243	0.955062	Polymyalgia Rheumatica	0.013213	Diseases of the musculoskeletal system and connective tissue
80	0.648763	0.977907	Epilepsy	0.016104	Diseases of the nervous system



### Comparison of APS/AUROC across Models

Figure S1, shows the performance three models, in terms of AUROC and APS, for each and every disease with a prevalence of 1% or higher. BEHRT's predictions remain in the upper right quadrant of the graph (for the most part) demonstrating a higher APS/AUROC than the other two models.



**Figure S1.** Disease-wise performance comparison for BEHRT, Deepr and RETAIN, all models trained on same dataset and trained on the same task (next 6 months), each point represents the APS/AUROC of one disease on corresponding model.

# **Gender-Specific Disease Analysis**

Table S7 illustrates the predictive performance of the model for gender-specific diseases. Male and female predictions in the table represent the number of patients with predictive probability above 0.5 for that gender, and disease gender indicates whether a disease is a male-specific or female-specific disease. M and F represent male and female respectively.

**Table S7.** Predictions of Gender-Specific Diseases (Test Predictions from Next 6 Months)

Disease Name	Disease Gender	Male Predictions	Female Predictions
Hyperplasia of Prostate	M	384	0
Hydrocoele (incl infected)	M	36	0
Male Infertility	M	1	24
Primary Malignancy Prostate	M	557	0
Erectile Dysfunction	M	425	1
Menorrhagia and Polymenorrhoea	F	0	697
Endometriosis	F	0	47
Female Genital Prolapse	F	0	865
Female Infertility	F	2	36
Benign Neoplasm of Ovary	F	0	69
Postmenopausal Bleeding	F	0	140
Primary Malignancy Breast	F	0	11
Primary Malignancy Ovarian	F	1	193

Male (M) and Female (F) Predictions imply label predictions >0.5



# Feature importance comparison

Table S8 illustrates an ablation study to selectively remove features (age, segment and position) to understand the importance of each feature.

**Table S8.** Feature Importance Comparison for the Next 6 Months Prediction Task

	F	eatures		Metrics		
Disease	Age	Segment	Position	APS	AUROC	
1	1	1	1	0.525	0.958	
1	0	1	1	0.515	$0.95^{2}$	
1	1	0	1	0.511	$0.95^{2}$	
1	1	1	0	0.498	0.954	
1	0	1	0	0.451	0.952	
1	1	0	0	0.501	0.955	
1	0	0	0	0.446	0.951	

1: Features activated, 0: Features deactivated

