# High-Performance Transformers for **Table Structure Recognition Need Early Convolutions**





ShengYun (Anthony) Peng

Seongmin Lee



Oral Presentation at NeurIPS'23 Workshop on Table Representation Learning







**Xiaojing** Wang **Rajarajeswari** Balasubramaniyan **Polo** Chau



# Existing table structure recognition research treats task as image-to-text generation

### Input Image

State/UT	Population			Number of literates			
	Total (millions)	Males (millions)	Females (millions)	Total literates (millions)	Males (millions)	Females (millions)	
J & K	10.1	5.4	4.8	4.8	3.1	1.7	
HP	6.0	3.1	3.0	4.0 14.8	2.3 8.4	1.8 6.3	
Punjab	24.3	13.0	11.4				
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3	
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1	
Haryana	21.1	11.4	9.8	12.1	7.5	4.6	
Delhi	13.9	7.6	6.2	9.7	5.7	3.0	
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7	

## Visual Encoder

# Textual Decoder



# Model architecture in existing method: Hybrid CNN-Transformer

- CNN backbone takes up ~50% of the total model parameters
- Significantly reduces both training and inference speed

#### Input Image

State/UT	Po	Population   fotal millions) Males (millions) Fer (millions)   10.1 5.4   6.0 3.1		Number of literates		
	Total (millions)	Males (millions)	Females (millions)	Total literates (millions)	Males (millions)	Females (millions)
J & K	10.1	5.4	4.8	4.8	3.1	1.7
HP	6.0	3.1	3.0	4.0	2.3	1.8
Punjab	24.3	13.0	11.4	14.8	8.4	6.3
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1
Haryana	21.1	11.4	9.8	12.1	7.5	4.6
Delhi	13.9	7.6	6.2	9.7	5.7	3.0
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7

#### Visual Encoder





Outputs (shifted right

Encoder &

Decoder

# But hybrid CNN-Transformers seldomly used: Vision Transformers use simple linear projection instead of CNN backbone



et. al.

# Can we simply employ the linear projection?

# **ResNet-18 CNN Backbone** LinearProj

State/UT	Pe	opulatio		Number of literates				
	Total (millions)	Males (millions	Females (millions)	T II li tates ( lions)	Males (millions)	Females (millions)		
J & K	10.1	5.4	4.0	4.8	3.1	1.7		
HP	6.0	3.1	3.0	4.0	2.3	1.8		
Punjab	24.3	13.0	11.4	14.8	8.4	6.3		
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3		
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1		
Haryana	21.1	11.4	9.8	12.1	7.5	4.6		
Delhi	13.9	7.6	6.2	9.7	5.7	3.0		
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7		

#### Input image

State/UT	Po	opulation		Number of literates			
	Total (millions)	Males (millions)	Females (millions)	Total literates (millions)	Males (millions)	Females (millions)	
J & K	10.1	5.4	4.8	4.8	3.1	1.7	
HP	6.0	3.1	3.0	4.0	2.3	1.8	
Punjab	24.3	13.0	11.4	14.8	8.4	6.3	
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3	
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1	
Haryana	21.1	11.4	9.8	12.1	7.5	4.6	
Delhi	13.9	7.6	6.2	9.7	5.7	3.0	
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7	

Model Performance Complexity High ✓ High X 28.70M #Param. 94.50% **Complex TEDS** 42.22G MAC Low X

86.62%

Lowest 22.67M 10.28G

Textual Decoder

**Visual Encoder Options for Table Structure Recognition** 



# **Can we simply employ the linear projection?** No, performance suffers

#### ResNet-18 CNN Backbone



### LinearProj

State/UT	Po	opulatio		Number of literates			
	Total (millions)	Males (millions	Females (millions)	T II li ates ( lions)	Males (millions)	Females (millions)	
J & K	10.1	5.4	4.0	4.8	3.1	1.7	
HP	6.0	3.1	3.0	4.0	2.3	1.8	
Punjab	24.3	13.0	11.4	14.8	8.4	6.3	
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3	
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1	
Haryana	21.1	11.4	9.8	12.1	7.5	4.6	
Delhi	13.9	7.6	6.2	9.7	5.7	3.0	
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7	

#### Input image

State/UT	Po	opulation	Number of literates			ates
	Total (millions)	Males (millions)	Females (millions)	Total literates (millions)	Males (millions)	Females (millions)
J & K	10.1	5.4	4.8	4.8	3.1	1.7
HP	6.0	3.1	3.0	4.0	2.3	1.8
Punjab	24.3	13.0	11.4	14.8	8.4	6.3
Chandigarh	0.9	0.5	4.0	0.6	3.8	0.3
Uttaranchal	8.5	4.3	4.2	5.1	3.0	2.1
Haryana	21.1	11.4	9.8	12.1	7.5	4.6
Delhi	13.9	7.6	6.2	9.7	5.7	3.0
Rajasthan	56.5	29.4	27.1	27.7	18.0	9.7

Performance

High ✓ 94.50% Complex TEDS Model Complexity

High X 28.70M #Param. 42.22G MAC

Low X 86.62% Lowest ✓ 22.67M 10.28G

Textual Decoder

Visual Encoder Options for Table Structure Recognition



# **Our Key Contribution & Discovery** ConvStem matches CNN backbone performance with a simpler model



56.5 29.4 27.1 27.7 18.0 9.7

Visual Encoder Options for Table Structure Recognition

erformance	Model Complexitv		
gh ✓ 1.50% mplex TEDS	High X 28.70M #Param. 42.22G MAC		
<b>w X</b> 5.62%	Lowest ✓ 22.67M 10.28G	Textual Decoder	
igh√ 4.66%	Low ✓ 24.08M 22.36G		

\_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_



## Why is convolutional stem effective? Higher receptive field ratio & longer sequence length



ConvStem matches CNN backbone performance with a simpler model



# **CNN Backbone** ResNet-34 has the highest TEDS due to its high RF ratio

Model	RF ratio (%)	Seq. length	TEDS (%)
ResNet-18	97.10	784	96.45
ResNet-34	100.00	784	96.76
ResNet-50	95.31	784	96.70



# **Linear Projection** As the patch size increases, performance generally improves, reaching its peak at a patch size of 56

Model	Patch size	RF ratio (%)	Seq. length	TEDS (%)	
LinearProj-112	112	25.00	16	90.61	Doak
LinearProj-56	56	12.50	64	92.17	
LinearProj-28	28	6.25	256	90.45	pertorma
LinearProj-16	16	3.57	784	87.56	
LinearProj-14	14	3.13	1024	87.22	

### Performance generally improves as patch size increases



# **Linear Projection** As the patch size increases, performance generally improves, reaching its peak at a patch size of 56

_	Model	Patch size	RF ratio (%)	Seq. length	TEDS (%)	
	LinearProj-112	112	25.00	16	90.61	Doak
	LinearProj-56	56	12.50	64	92.17	r Car
	LinearProj-28	28	6.25	256	90.45	performa
	LinearProj-16	16	3.57	784	87.56	
	LinearProj-14	14	3.13	1024	87.22	

#### Patch size too large - sequence length too small - worse performance (Patch size & sequence length inversely correlated)





# **Convolutional Stem** Optimal balance of RF ratio & sequence length

Model	RF ratio (%)	Seq. length	TEDS (%)	
ConvStem-R3 ConvStem-R2 ConvStem-R1	13.62 12.95 12.82 6.92	729 729 784 784	96.53 96.02 96.14 95.57	Higher RF ratio increases TEDS
ConvStem-N3 ConvStem-N2 ConvStem-N1	12.10 13.62 15.56 12.30	900 729 528 256	96.50 96.53 95.89 94.32	Longer sequence length increase TEDS





Presenting complaints	N (% of total sample n = 259288)	*Pulse	\$	۲۲ <b>۲</b>	*603	*RR	"\$p02
			N (% (	of each presen	tin <mark>g co</mark> mpla	aints)	
Fever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)
Injury (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (15)	2267 (7.6)	1437 (4.8)
Abdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5)	311 (13)	1520 (66)	732 (3.2)
Chest pain	20130 (7.8)	8630 (42.9)	90% (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (7.0)
Injury (Head/face/heoli)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (64)
Vomiting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)
Headache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	286 (3.0)	887 (93)	271 (28)
Shortness of breath	8548 [3.3]	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)
Back pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (O.S)	368 (45)	169 (2.1)
Diarrhea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)
Total	160744 (62.0)	41521 (258)	43649 (27.8)	38507 (24.0)	4550 (2.8)	18556 (11.5)	7619 (4.7)

<b>√</b>	<th>r&gt;</th> <th></th> <th></th> <th></th> <th></th>	r>				
	*Pulse	"BP	* T*C	*GCS	188	"SpO2

	(% or total sample n = £33286)							
		N (% of each presenting complaints)						
ſ	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)	
y (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (1 <i>5</i> )	2267 (7.6)	1437 (4.8)	
ominal pain	23170 (8.9)	5885 (25.4)	5986 (25.8)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)	
t pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (68)	5157 (25.6)	1409 (7.0)	
y (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (64)	
iting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)	
tache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	296 (3.0)	887 (93)	271 (28)	
tness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)	
pain	8239 (3.2)	1557 (189)	1997 (19.4)	1392 (16.9)	39 (0.5)	368 (45)	169 (2.1)	
hea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)	
	160744 (62.0)	41521 (25.8)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)	

Presenting complaints	N (% of total sample n = 259288)	*Pulse	192	۲°C	'6CS	*RR	"SpO2		
		N (% of each presenting complaints)							
Fever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)		
lnjuty (non-head/face/heck)	29695 (11.4)	2767 (93)	3200 (10.8)	2583 (8.7)	441 (15)	2267 (7.6)	1437 (4.8)		
Abdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5))	311 (13)	1520 (66)	732 (3.2)		
Chest pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (7.0)		
Injury (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)		
Vomiting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)		
Headache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	286 (3.0)	887 (9.3)	271 (28)		
Shortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (7.2)		
Back pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (45)	169 (2.1)		
Diarrhea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)		
Total	160744 (62.0)	41521 (258)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)		

**√** 

Presenting complaints	N (% of total sample n = 259288)	*Pulse	\$	• TC	*GCS	*RR	"SpO2	
		N (% of each presenting complaints)						
Fever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (8.7)	3797 (12.0)	1288 (4.1)	
injury (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (8.7)	441 (15)	2267 (7.6)	1437 (4.8)	
Abdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)	
Chest pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (7.0)	
Injury (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)	
Vomiting	10629 (4.1)	3135 (295)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)	
Headache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	286 (3.0)	887 (93)	271 (28)	
Shortness of breath	8548 (3.3)	2711 (81.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (7.2)	
Back pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (4.5)	169 (2.1)	
Diarrhea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)	
Total	160744 (62.0)	41521 (25.8)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)	

	2310 [32]	2201 (343)	21/00 (20/2)
hortness of breath	8548 [3.3]	2711 (31.7)	3087 (36.1)
Back pain	8239 (3.2)	1557 (189)	1997 (19.4)
Dianthea	5954 (2.3)	1835 (30.8)	1916 (32.2)
lotal	160744 (62.0)	41521 (25.8)	43649 (27.8)
		-	
		∕tr	
		<tr< th=""><th>&gt;</th></tr<>	>
Presenting complaints	N (% of total sample n = 259288)	<tr< th=""><th>₩ ₩</th></tr<>	₩ ₩
Presenting complaints	N (% of total sample n = 259288)	<tr <sup>*Pulse</sup></tr 	> *#P N (% o
Presenting complaints	N (% of total sample n = 259288) 31554 (122)	<tr "Pulse 9964 (31.6)</tr 	*8P N (% o 9645 (30.6)
Presenting complaints ever nury (non-head/face/heol)	N (% of total sample n = 259288) 31554 (12.2) 29095 (11.4)	*Pulse   9964 (\$1.6)   2267 (\$3.3)	> *8P N (% of 9645 (30.6) 3200 (10.8)
Presenting complaints ever njury (non-headiface/head) Nadominal pain	N (% of total sample n = 259288) 31554 (12.2) 29095 (11.4) 23170 (8.9)	*Pulse   9964 (31.6)   22/67 (9.3)   5886 (25.4)	> *BP N (% o 9645 (30.6) 3200 (10.8) 5996 (25.8)
Presenting complaints Fever nyury (non-head/face/heck) Andominal pain Chest pain	N (% of total sample n = 259288) 31554 (122) 29095 (11.4) 23170 (8.9) 20130 (7.8)	"Pulse   9964 (31.6)   2267 (9.3)   5886 (25.4)   8630 (42.9)	*8P       N (% of       9645 (30.6)       3200 (10.8)       5996 (25.8)       9076 (45.1)

senting complaints	N	*Pulse	18P	<b>۲</b> ۲	'GCS	*RR	"5p02
	(% of total sample n = 259288)						
			N (% c	of each presen	ting compla	uints)	
er	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)
ny (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (1.5)	2267 (7.6)	1437 (4.8)
dominal pain	23170 (8.9)	5895 (25.4)	5986 (25.8)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)
est pain	20130 (7.8)	8630 (42.9)	90% (45.1)	7008 (34.8)	1371 (68)	5157 (25.6)	1409 (7.0)
ny (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (64)
miting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)
adache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	296 (3.0)	887 (93)	271 (28)
ortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)
k pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (45)	169 (2.1)
rthea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)
al	160744 (62.0)	41521 (25.8)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)

#### Step 68

(% of total sample n = 259288)		ruse	.DL	11	.00	·M	.shot	FIESE		
	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)	Fever		
head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (1.5)	2267 (7.6)	1437 (4.8)	Injury (r		
pain	23170 (8.9)	5895 (25.4)	5986 (25.8)	4516 (19.5))	311 (13)	1520 (66)	732 (3.2)	Abdom		
	20130 (7.8)	8630 (42.9)	90% (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (7.0)	Chest p		
d/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)	Injury (F		
	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)	Vomitin		
	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	296 (3.0)	887 (93)	271 (2.8)	Headac		
of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (7.2)	Shortne		
	8239 (3.2)	1557 (189)	1997 (194)	1392 (16.9)	39 (O.S)	368 (45)	169 (2.1)	Back pa		
	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)	Diantes		
	160744	41521 (25.8)	43649 (27.8)	38507	4550	18556	7619	Total		

റ	cr	າລເ	n =	"6"
CO	S	Jai	1-	0

Presenting complaints	N (% of total sample n = 259288)	*Pulse	фр	* 110	'GCS	*RR	"\$p02	
		N (% of each presenting complaints)						
Fever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (87)	3797 (12.0)	1288 (4.1)	
Injury (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (1.5)	2267 (7.6)	1437 (4.8)	
Abdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)	
Chest pain	20130 (7.8)	8630 (42.9)	90% (45.1)	7008 (34.8)	1371 (6.8)	5157 (ZS.6)	1409 (7 <i>0</i> )	
Injury (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)	
Vomiting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)	
Headache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	285 (3.0)	887 (93)	271 (2.8)	
Shortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)	
Back pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (D.S)	368 (4.5)	169 (2.1)	
Diarthea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)	
Total	160744 (62.0)	41521 (258)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)	

Step 26

resenting complaints	N (% of total sample n = 259288)	"Pulse	'BP	11	°65	*KK	"SpO3	
		N (% of each presenting complaints)						
ever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1	
njury (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (15)	2267 (7.6)	1437 (48	
ibdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.B)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)	
hest pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (71	
njury (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)	
lomiting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)	
leadache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	286 (3.0)	887 (93)	271 (2.8)	
hortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)	
lack pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (45)	169 (2.1)	
Xanhea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1 <i>8</i> )	608 (10.2)	244 (4.1)	
otal	160744	41521 (258)	43649 (27.8)	38507	4550	18556	7619	

**X** 

**√**

2.8 (11.5) (4.7)

senting complaints	N (% of total sample n = 259288)	*Pulse	÷	* T'C	*605	*RR	"SpO2			
		N (% of each presenting compla					ints)			
er	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)			
ny (non-head/face/heck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (87)	441 (1.5)	2267 (7.6)	1437 (48)			
dominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5))	311 (1.3)	1520 (66)	732 (3.2)			
est pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (68)	5157 (25.6)	1409 (7.0)			
ny (Head/face/heck)	13309 (5.0)	1769 (133)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (64)			
miting	10629 (4.1)	3135 (295)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)			
adache	9516 (3.7)	3267 (343)	3706 (38.9)	2779 (29.2)	296 (3.0)	887 (93)	271 (28)			
ortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (72)			
k pain	8239 (3.2)	1557 (189)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (45)	169 (2.1)			
rthea	\$954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)			
al	160744 (62.0)	41521 (258)	43649 (27.8)	38507 (240)	4550 (2.8)	18556 (11.5)	7619 (4.7)			

Step 67

rowspan="2"

oss-attention Maps

### **CNN Backbone** (ResNet-18) Miscounts columns in header spanning cell TEDS: 98.28%

### Linear Projection (LinearProj-28) Mis-predict html tags due to scattered attention TEDS: 86.07%

## **ConvStem**

Accurately reconstructs table TEDS: 100.00%



# Easy-to-Use Open-Source Research github.com/poloclub/tsr-convstem

## Step 1. Configure experiment

## Step 2. Train & evaluate the model

\$ make experiments/r18\_e2\_d4\_adamw/.done\_teds\_structure



#### EXP\_r18\_e2\_d4\_adamw := \$(PUBTABNET) \$(MODEL\_r18\_e2\_d4) \$(OPT\_adamw)



github.com/poloclub/tsr-convstem

# Thanks **High-Performance Transformers** for Table Structure Recognition **Need Early Convolutions**

#### **ConvStem matches classic CNN backbone performance, with a much simpler model.**



**Visual Encoder Options for Table Structure Recognition** 







Textual

Decoder



Lee



Rajarajeswari Balasubramaniyan







Polo Chau

