A Appendices

A.1 Implementation Details

Our classifier for INDICATION is a two-layer network, in which the first layer has the size of 50 and 5 for the second layer. ReLU activation and dropout are used in-between the two layers. The dropout rate is set to be 0.1. We also use a two-layer network for the QUANTITY regression task, in which the first layer has the size of 10 and its second layer predicts the value for *quantity*. We also add a dropout the rate 0.1 in-between the two layers. Our classifier for QUANTITY TAG has only one layer the size of 5.

We reduce the dimension of attended vector before signature transform from 768 to 32, as depicted in Figure 2 (left). We set the order of truncated signature to be 2, and employ 8 parallel attention layers (or heads). We ran grid-search over two choices of learning rate $lr : [0.00003, 0.00005]$, and 10 randomly sampled triplets of α_{qnt}, β_{qntt} and β_{ind}. Each value of α_{qnt}, β_{qntt} and β_{ind} is between 0 and 1 respectively, while we make sure each combination of the three values sums to 1.

A.2 Signature of paths

We begin with the definition of the signature, using more traditional notation of stochastic calculus.

Definition 1. *Let $X = (X^1, ..., X^d)$ be a path in $\mathbb{R}^d$. The signature of X is defined as the infinite collection of iterated integrals:*

$$S(X) = \left(\int ... \int_{a<t_1<...<t_k<b} dX_{t_1} \otimes ... \otimes dX_{t_k} \right)_{k \geq 0}$$
$$= \left(\left(\int ... \int_{a<t_1<...<t_k<b} dX_{t_1}^{i_1} ... dX_{t_k}^{i_k} \right)_{1 \leq i_1,...,i_k \leq d} \right)_{k \geq 0}$$

where $dX_t = \frac{dX_t}{dt} dt$ and the $k = 0$ term is taken to be $1 \in \mathbb{R}$.

Definition 2. *The truncated signature of order N of X is defined as:*

$$S^N(X) = \left(\int ... \int_{a<t_1<...<t_k<b} dX_{t_1} \otimes ... \otimes dX_{t_k} \right)_{0 \leq k \leq N}.$$

The dimension of the truncated signature explodes exponentially with the input path dimension:

Proposition 1. *For any $d \geq 1$, the truncated signature of order N of a d-dimensional path has the dimension of:*

$$\sum_{k=0}^{N} d^k = \frac{d^{N+1} - 1}{d - 1}$$

In practice, the term of order 0 is dropped as it is always equal to 1. For clarity, we define:

$$S(X)^{i_1,...,i_k} = \int ... \int_{a<t_1<...<t_k<b} dX_{t_1}^{i_1} ... dX_{t_k}^{i_k}$$

with $1 \leq i_1, ..., i_k \leq d$, so that:

$$S(X) = \left(\left(S(X)^{i_1,...,i_k} \right)_{1 \leq i_1,...,i_k \leq d} \right)_{k \geq 0}$$
$$= (1, S(X)^1, ..., S(X)^d, S(X)^{1,1}, S(X)^{1,2}, ...).$$

d_{presig}	order_{sig}	d_{sig}
512	1	512
512	2	262K
256	2	66K
128	2	16K
128	3	2M
64	2	4K
64	3	266K
32	2	1057
32	3	34K
16	2	272

d_{presig}	order_{sig}	d_{sig}
16	3	4K
8	2	72
8	3	584
8	4	5K
4	4	340
4	5	1365
4	6	5K
2	9	1022
2	10	2K
2	12	8K

Table 5: The number of dimensions (d_{sig}) of the truncated signature is determined by the size of its input (d_{presig}) and the order of truncation selected (order_{sig}).

Model	**QUANTITY TAG**					**QUANTITY**
	Standard	**APPP**	**PRN**	**Complex**	**NS**	
Base	0.71	0.50	0.10	0.76	0.95	0.50
ClinicalBERT	0.83	0.97	0.89	0.99	0.79	0.21
M-BERT	0.94	0.41	0.04	0.81	0.98	0.41
Base + LSTM	0.93	0.36	0.99	0.77	0.74	0.45
ClinicalBERT + LSTM	0.90	0.23	0.99	0.61	0.49	0.47
M-BERT + LSTM	0.29	0.99	0.72	0.47	0.93	0.50
Base + STE	0.97	0.84	0.77	0.44	0.88	0.23
ClinicalBERT + STE	0.99	0.70	0.85	0.65	0.65	0.36
M-BERT + STE	0.86	0.97	1.00	0.89	0.86	0.15

Table 6: Model performance comparison for QUANTITY and also across different classes in QUANTITY TAG. APPP: As Per Previous Prescription; NS: Not Specified.

Model	**INDICATION**				
	Cardiac	**Tremors**	**Migraine**	**Others**	**NA**
Base	0.03	0.05	0.02	0.27	0.03
ClinicalBERT	0.00	0.24	0.03	0.04	0.13
M-BERT	0.05	0.01	0.09	0.35	0.00
Base + LSTM	0.02	0.00	0.00	0.31	0.00
ClinicalBERT + LSTM	0.01	0.03	0.34	0.08	0.02
M-BERT + LSTM	0.05	0.00	0.00	0.04	0.03
Base + STE	0.00	0.00	0.02	0.08	0.05
ClinicalBERT + STE	0.00	0.26	0.05	0.02	0.00
M-BERT + STE	0.00	0.11	0.00	0.00	0.15

Table 7: Model performance comparison across different classes in INDICATION. NA: Not Annotated.

Indication (44)	
CARDIAC - HYPERTENSION - PALPITATIONS	49
CARDIAC - PALPITATIONS - TREMORS	1
CARDIAC - PALPITATIONS - ANGINA	1
CARDIAC - PALPITATIONS - ATRIAL FIBRILLATION	1
CARDIAC - ATRIAL FIBRILLATION	57
CARDIAC - DYSRHYTHMIA	101
CARDIAC - HEART FAILURE	60
CARDIAC - HYPERTENSION - ANGINA - PALPITATIONS	1
CARDIAC - ATRIAL FIBRILLATION - DYSRHYTHMIA/RATE	1
CARDIAC - HYPERTENSION - DYSRHYTHMIA	19
CARDIAC - MIGRAINE	1
CARDIAC - DYSRHYTHMIA/RATE	114
CARDIAC - HYPERTENSION - ESOPHAGEAL VARICES	1
CARDIAC - PALPITATIONS	162
CARDIAC - HYPERTENSION - HEART FAILURE	4
CARDIAC - HYPERTENSION - MIGRAINE	12
CARDIAC - HYPERTENSION - ANGINA	28
CARDIAC - TREMORS	2
CARDIAC - DYSRHYTHMIA - MIGRAINE	1
CARDIAC - ANGINA - ATRIAL FIBRILLATION	1
CARDIAC - HEPATIC CIRCULATION	1
CARDIAC - ANGINA	34
CARDIAC - DYSRHYTHMIA/RATE - PALPITATIONS	1
CARDIAC - HYPERTENSION	1655
CARDIAC - HYPERTENSION - ATRIAL FIBRILLATION	2
CARDIAC - ESOPHAGEAL VARICES	2
CARDIAC	608
CARDIAC - ANGINA - DYSRHYTHMIA/RATE	1
CARDIAC - HYPERTENSION - DYSRHYTHMIA/RATE	50
CARDIAC - PALPITATIONS - MIGRAINE	1
CARDIAC - HYPERTENSION - TREMORS	8
ESOPHAGEAL VARICES	1
ANXIETY - CARDIAC - PALPITATIONS	3
ANXIETY	6
ANXIETY - MIGRAINE	1
HYPERTHYROIDISM	1
HYPERGLYCEMIA	2
CHEST PAIN	1
TREMORS	77
TREMORS - CARDIAC - HYPERTENSION	1
TREMORS - ANXIETY	1
TREMORS - CARDIAC - PALPITATIONS	2
MIGRAINE	69
NA	707

Figure 5: Number of prescriptions per indication class, where the indication label has the original 44 classes. *NA* stands for *Not Annotated*.

Swedish	Translated English	Indication	Quantity	Quantity Tag
0.5 TABLETTER 2 GANGER DAGLIGEN. FOR HJARTRYTMEN OCH SANKER BLODTRYCKET TABLETTERNA SVALJES HELA (KAN DELAS VID SVALJSVARIGHETER, MEN FAR EJ K	0.5 tablets 2 times daily. For the heart rhythm and reduces blood pressure The tablets are swallowed whole (can be divided at swallowing durations, but don't k	Cardiac-hypertension	1	Standard
1 TABLETT 1 GANG DAGLIGEN I 3 VECKOR. EVENTUELL HOJNING TILL 2 TABLETTER DAGLIGEN BEROENDE PA KONTROLLEN OCH SYMTOMLINDRING, MOT HJARTKLAPPNING	1 table 1 time daily for 3 weeks. Successful increase for 2 tablets daily dependent on control and symptoms relief, against palpitations	Cardiac-palpitations	1	Complex
1 TABLETT 1 GANG DAGLIGEN. EFTERHAND EVENTUELL OKNING TILL EN TABLETT MORGON OCH LUNCH MEN BORJA MED EN TABLETT TIDIG MORGON, MOT SKAKNINGAR.	1 tbale 1 time daily. Previously opening to a tablet morning and lunch but begin with a tablet early morning, against shakes	NA	1	Standard
1/2 TABLETT PA MORGONEN , 1/2 TABLETT PA LUNCHEN OCH 1/2 TABLETT PA KVALLEN MOT TREMOR, BLODTRYCKSREGLERANDE	1/2 table on morning, 1/2 table on lunch and 1/2 table on evening against tremor, blood pressure control	NA	1.5	Standard
1 TABLETT 1 GANG DAGLIGEN MOT HOGT BLODTRYCK DU BOR BESTALLA LAKARTIDLAMPLIGEN MAJ JUNI FOR KONTROLL BT DIABETES.	1 tablet 1 time daily against high blood pressure you should order the painting lighting may june for control BT diabetes	Cardiac-hypertension	1	Standard
2 depottabletter kl. 08, 1 depottablett kl. 20. Dagligen. Mot hOgt blodtryck	2 prolonged-release tablets at. 08, 1 prolonged-release tablet at. 20. Daily. Against high blood pressure	Cardiac-hypertension	3	Standard

Table 8: Longer example prescriptions with translations and annotations.