

ICD10 chapter 2:
Neoplasms

Test prevalence: 17.2%

Macro ROC AUC

0.8
0.7
0.6
0.5
0.4

0

1000

2000

3000

4000

5000

6000

7000

8000

Number of patients in train set

Featurizers

Train Embeddings

SNDS Embeddings

Cui2vec Embeddings

Count Encoding

Count Encoding wo decay

CEHR-BERT

Estimators

Logistic regression

Forests

