

ICD10 chapter 13:
Diseases of the
musculoskeletal system and
connective tissue

Test prevalence: 11.6%

Macro ROC AUC

0.60
0.58
0.56
0.54
0.52
0.50
0.48
0.46

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

