Comparing deep-learning and concept extraction based methods for patient phenotyping from the clinical narrative

Deep Shah (deepss2)
Sandeep yerramsetty (sy40)

Course Project
UIUC CS598 - Spring 2023
Paper Link

Outline

- What is patient phenotyping
- Introduction to the paper
- What we reproduced
- Claim 1: CNN outperform N-gram
- Claim 2: CNN with width 1-2 outperform width 1 CNN
- Ablation study motivation
  - Importance of CNN part
  - Importance of MaxPool

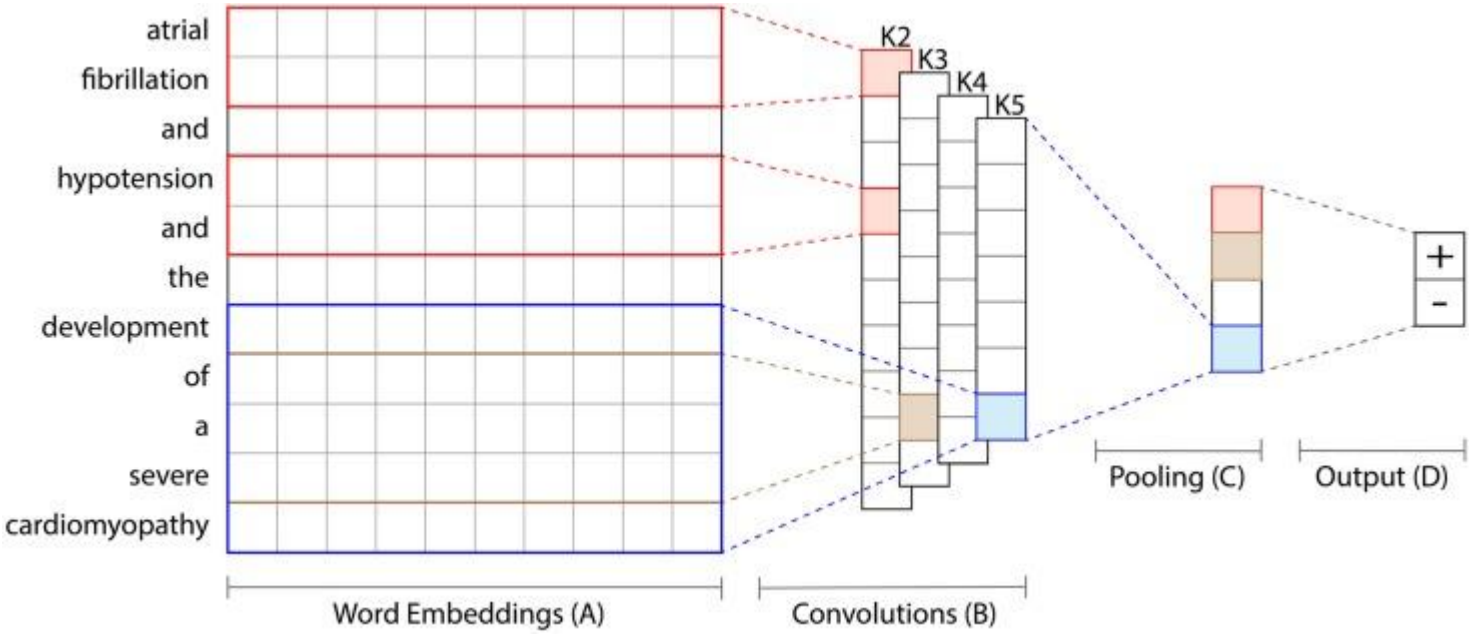
What is Patient Phenotyping

- A classification task that classifies whether the patient has a particular condition or is likely to develop one in the future based on the Electronic health records(EHR) data
- Applications for patient phenotyping can support clinicians by reducing the time they spend on chart reviews, which takes up a significant fraction of their daily workflow

Introduction to the paper

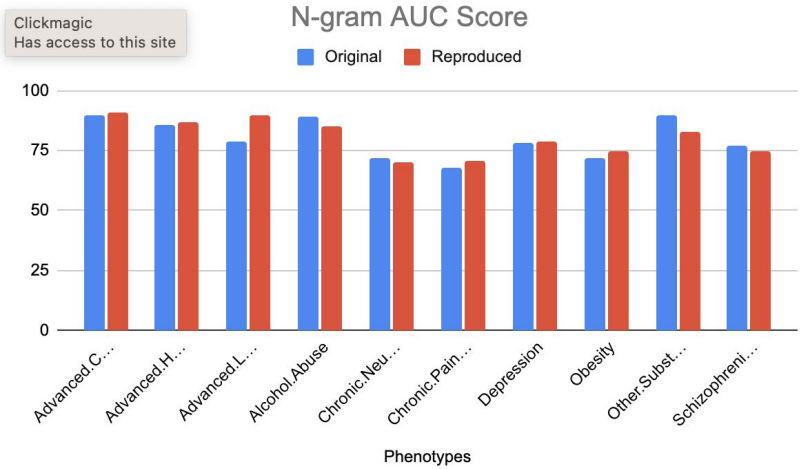
- The paper compares concept extraction based methods with CNNs and other commonly used models in NLP in ten phenotyping tasks using 1,610 discharge summaries from the MIMIC-III database.
- We show that CNN's outperform concept extraction method and other baseline models in almost all of the tasks, with an improvement observed in F1 and AUC (Area under curve) score.
- The results indicate that CNNs are a valid alternative to existing approaches in patient phenotyping and cohort identification, and should be further investigated
- The deep learning approach presented in this paper can be used to assist clinicians during chart review or support the extraction of billing codes from text by identifying and highlighting relevant phrases for various medical conditions

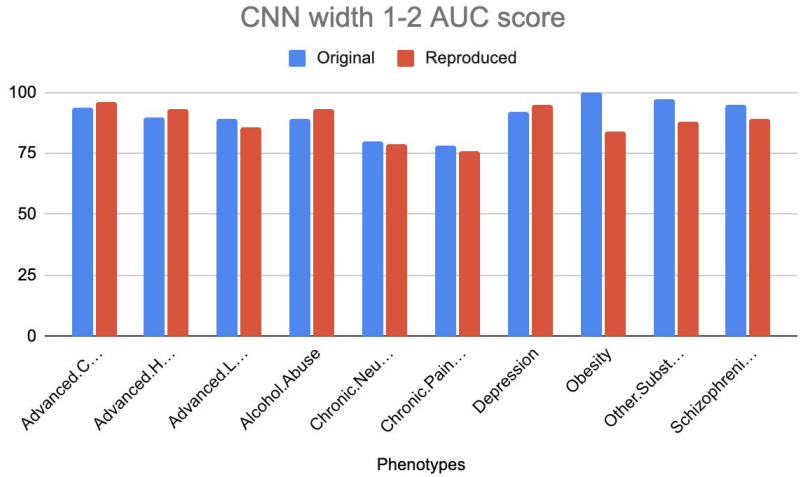
Model architecture



What we reproduced

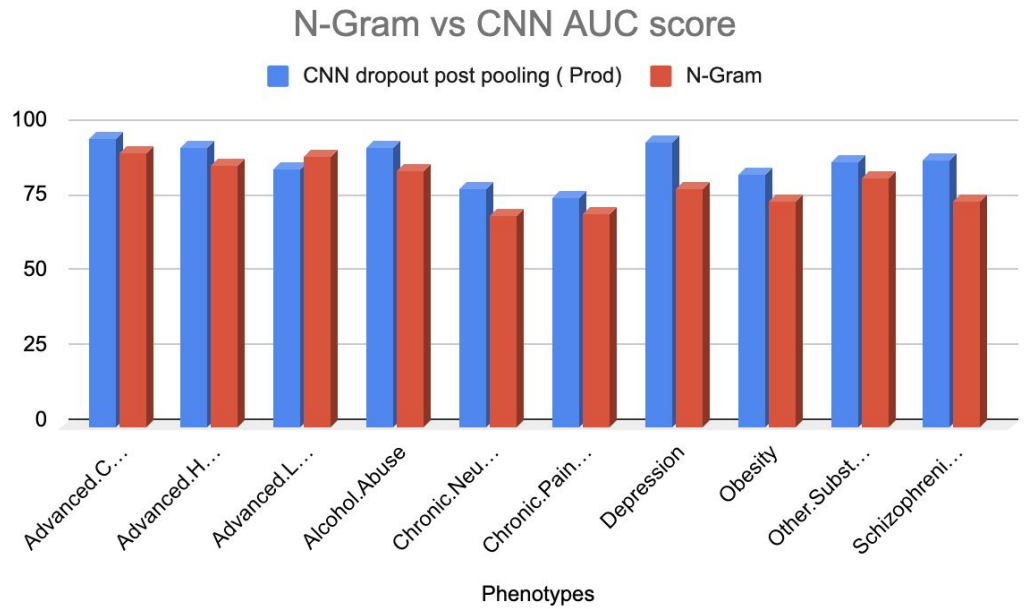
- We reproduced the word2vec embedding based on discharge summaries.
- We did reproduce the N-gram with various N's.
- We reproduced the architecture proposed by the author for various CNN widths and used the learnt word2vec embeddings.
- Results are matching what the author reported.





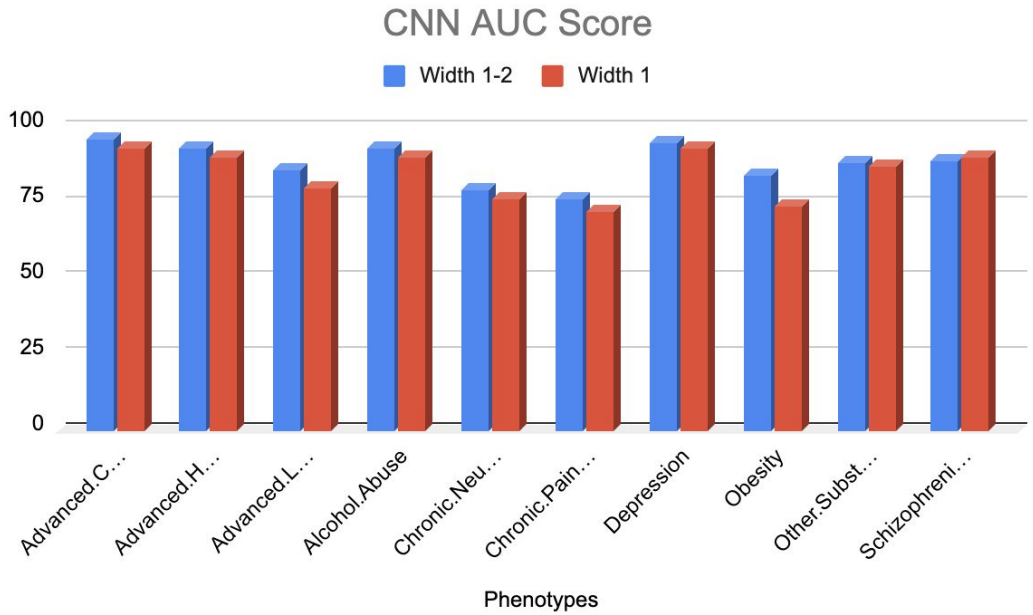
Claim 1: CNN outperform N-gram

- We were able to reproduce both the N-gram and CNN based model.
- Our reproduced study also support the hypothesis that that the CNN model outperforms the N-gram based model.



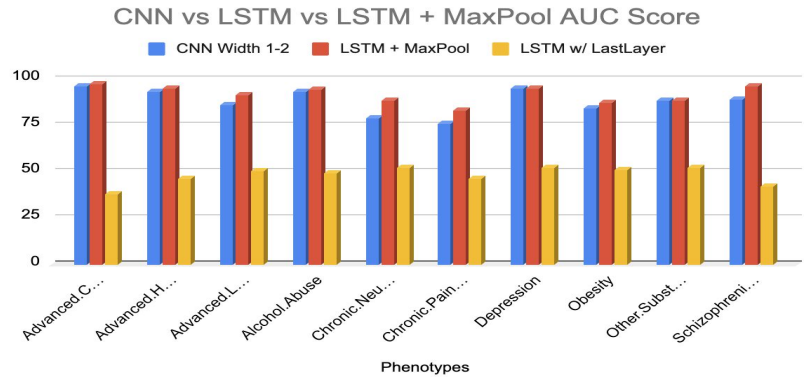
Claim 2: CNN with width 1-2 outperform width 1 CNN

- We can see that the the CNN with width 1-2 have much better AUC score compared to CNN with width 1.



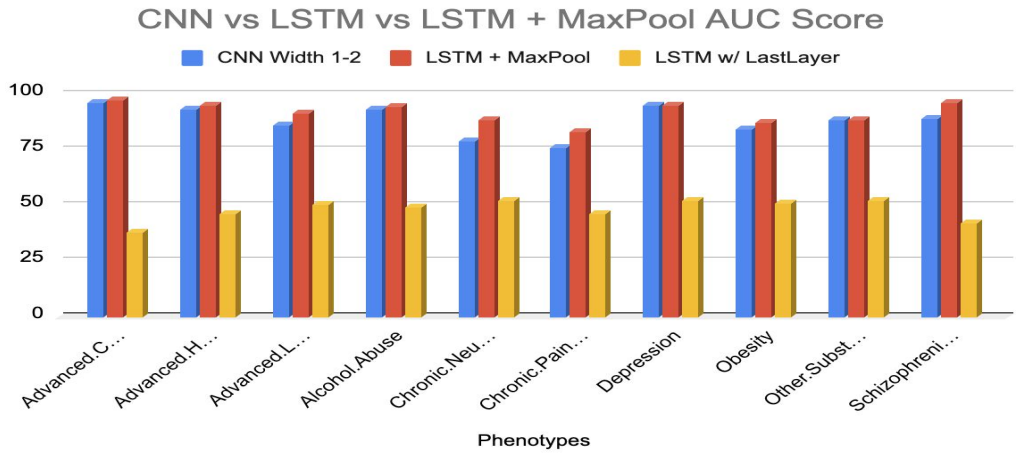
Ablation: Import of CNN in architecture

- Given the advancement in NLP based technique for Deep learning, we decided to use replace CNN with LSTM.
- The output embedding from each token will be passed to MaxPool similar to the original model.
- The model (Red) performs as good as the CNN model (Blue), and beating CNN based model on certain phenotypes.



Ablation: Importance of MaxPool

- To measure the importance of MaxPool on our data, we decided to use the last layer of LSTM to see whether it can learn the important concepts of phenotype.
- The AUC score (Yellow) is quite less compared to CNN + MaxPool (Blue), which suggests that MaxPool does play important role in our model.



Thank you