Predicting Heart Failure Readmission from Clinical Notes Using Deep Learning

By Lamia Abbas & Nelofer Arjumand

General Problem

- Heart failure is a prevalent issue that leads to hospital readmission
  - High financial and resource costs
  - Burden on patients
- Predicting readmission risk
  - Early diagnosis
  - Prevent readmissions
- Previously proposed models
  - Use structured clinical data
  - Require extensive feature engineering

What the specific approach taken was

- Leverages pre-trained word embeddings to represent words in the clinical notes
  - Word2Vec pretrained on PubMed and PubMed Central texts
- CNN architecture automatically generates feature maps for training
  - Replaces the need of feature engineering
- Train models to predict whether an admission is likely to be followed by a readmission.
- Conduct experiments involving prediction of two outcomes: general readmission and 30-day readmission.
- Compared results
  - Word2Vec + CNN versus TF-IDF + Random Forest

What the results claimed were

Original paper 30-day prediction task:

	precision	recall	f1-score	accuracy
Deep learning (CNN)	0.698	0.771	0.733	0.7188
Random forest (TF-IDF)	0.690	0.625	0.656	0.6719

- Combination of NLP and Deep Learning outperform conventional machine learning models for prediction tasks
- Specifically, the CNN model using pretrained Word2Vec embeddings outperforms a baseline Random Forest Classifier with Term Frequency - Inverse Document Frequency (TF-IDF) in the 30-readmission prediction task.

Our reproduction attempts

- Re-created dataset from MIMIC-III clinical notes, admissions, and diagnosis codes
  - Narrowed down only admissions with a heart failure diagnosis code, designated by ICD-9 codes: 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, and 428.9
  - Chose admissions with 30-day readmissions
  - Chose clinical notes that were discharge summaries
- Clinical notes are converted to embeddings where each word is represented by a word vector
  - Loaded pre-trained Word2Vec embeddings
- Then input to our CNN model; 3 convolutional layers with ReLU activation, followed by max-pooling and dropout layers

Our results

Our reproduced 30-day readmission prediction task:

	precision	recall	f1-score	accuracy
Deep learning (CNN)	0.513811	0.543746	0.473926	0.675145
Random forest (TF-IDF)	0.537575	0.634859	0.472003	0.613372

- Our results supported the claim that CNN with Word2Vec embeddings outperform the random forest with tf-idf weights
- CNN has a higher accuracy than RF, although f1-scores are similar

Dataset Preprocessing

	Positive samples (admissions with 30-day readmissions)	Negative samples (admissions without 30-day readmissions)	Total Samples	Prevalence of positive samples
Imbalanced training data	1,044	11,335	12,379	8.43%
Training data after undersampling	1,044	1,044	2,088	50%
Training data with augmentation	2,044	11,335	13,379	15.27%
Training data with augmentation/undersampling	2,044	2,044	4,088	50%

Resources & References

Link to github repo:
https://github.com/nelarj2/DeepLearningForHealthCare_FinalProject

Reference: Xiong Liu, Yu Chen, Jay Bae, Hu Li, Joseph Johnston, and Todd Sanger. 2019. Predicting heart failure readmission from clinical notes using deep learning.