Project Proposal

Rezana Ganie & David Alejandro Henriquez
rganie3@gatech.edu & dhenriquez3@gatech.edu

Paper 1: index: 41, paper title: AI-Driven Clinical Decision Support: Enhancing Disease Diagnosis Exploiting Patients Similarity, venue: IEEE Access, authors: Carmela Comito, Deborah Falcone, and Agostino Forestiero
Task: The task proposes a CDS (Clinical Decision Support) framework which is used to provide patient diagnosis suggestions to physicians.
This achieved by extracting characteristics of patients from disparate sources. The approach uses word embeddings that can model the semantic relationships which can measure the different patient diagnosis. This is achieved through analyzing symptoms similarity in order to make a prediction.
Innovation: The innovation of this research is to enhance existing CDS systems. For example the current solution focuses on a single patient at time while ignoring other similar patient.
Disadvantages/Advantages: The advantage was that this model performed on a real- world EHR. The disadvantages of this research limits patient diagnosis to historically similar patients
Data Accessibility: Yes. MIMIC III v1.4, a publicly available critical care dataset. http://staff.icar.cnr.it/diseaseDiagnosis.zip
Code Accessibility: Yes, http://staff.icar.cnr.it/diseaseDiagnosis.zip

Paper 2: index: 201, paper title: FarSight: Long-Term Disease Prediction Using Unstructured Clinical Nursing Notes, venue: IEEE Transactions on Emerging Topics in Computing, authors: Tushaar Gangavarapu, Gokul S Krishnan, Sowmya Kamath S, Jaykumar Jeganathan
Task: Build a model that recognizes the onset of a disease by using the earliest detected symptoms. The principal source of data is nurse clinical notes. The model uses long-term aggregation systems before the NLP tasks.
Innovation: The use of unstructured clinical notes and not just EHR data. A long-term aggregation system.
Disadvantages/Advantages: The model beats state-of-the-art EHRs models. However, it is biased toward nursing data (hard to generalize to different sources of data). Don't give many insights on how they implemented the first step aggregation model Farsighted (long-term aggregation by future like up).

Data Accessibility: Yes. MIMIC III, a publicly available critical care dataset. **Code Accessibility:** Code is not provided by the author.

Paper 3: index: 235, paper title: Med7: A transferable clinical natural language processing model for electronic health records, venue: Artificial Intelligence in Medicine, authors: Andrey Kormilitzin, Nemanja Vaci, Qiang Liu, Alejo Nevado-Holgado
Task: An NLP model that recognizes drug names, routes of administration, frequency, dosage, strength, form, and duration from clinical data. More precisely a named-entity recognition (NER) model for clinical natural language processing with good transferability properties. The NER task seeks to classify words in to the seven predefined categories mentioned above. The paper extracts information from free-text.
Innovation: A model that adapts to a different dataset with only a small fine-tuning on a small sample dataset. The model is trained with MIMIC III data and transferred with transfer learning to UK CRIS dataset.
Disadvantages/Advantages: The model is poses difficulty when generalizing. The one advantage to this the model presents a high score overall categories after fine-tuning on a small sample the transfer model has a good performance.
Data Accessibility: Yes, MIMIC-III data and UK-CRIS network data.
Code Accessibility: Yes, https://github.com/kormilitzin/med7

Decide on your target paper:

Paper 1, AI-Driven Clinical Decision Support: Enhancing Disease Diagnosis Exploiting Patients Similarity. We selected this Paper first because it applies concepts taught in class, such as handling health care data, patient similarity, and patient diagnosis. We are interested in learning more about NLP. Hypotheses – A new CDS framework that introduces a mechanism to measure the semantic relationship of different diagnoses in terms of symptoms similarity for the prediction produces good precision and recall scores when providing patient diagnosis as an alternative solution to the existing CDS.

The data and code is available on http://staff.icar.cnr.it/diseaseDiagnosis.zip. The software requirements public libraries. We will use python for our solution, for the hardware we intend on using a virtual machine with HPC however if we cannot do this, we will use our local machines and process the data in batches.

References

1. Comito, C., Falcone, D., & Forestiero, A. (2022). AI-driven clinical decision support: Enhancing disease diagnosis exploiting patients similarity. IEEE Access, 10, 6878–6888. https://doi.org/10.1109/access.2022.3142100
2. Gangavarapu, T., Krishnan, G. S., S, S. K., & Jeganathan, J. (2021). Farsight: Long-term disease prediction using unstructured clinical nursing notes. IEEE Transactions on Emerging Topics in Computing, 9(3), 1151–1169. https://doi.org/10.1109/tetc.2020.2975251
3. Kormilitzin, A., Vaci, N., Liu, Q., & Nevado-Holgado, A. (2021). Med7: A transferable clinical natural language processing model for Electronic Health Records. Artificial Intelligence in Medicine, 118, 102086. https://doi.org/10.1016/j.artmed.2021.102086