Project Proposal

Rezana Ganie & David Alejandro Henriquez  
[rganie3@gatech.edu](mailto:rganie3@gatech.edu) & [dhenriquez3@gatech.edu](mailto: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

1. 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 such as lab results, medical image(s) or electronic health records.   
     
   The approach uses word embeddings that can model the semantic relations between (1)hospital admissions, (2)symptoms and (3)diagnosis. The approach introduces a tool which can measure the relationships of different diagnosis.

This is achieved through analyzing symptoms similarity in order to make a prediction. The semantic models are created from all collected data and an AI method is developed to generate context-based and rich representation of health related information.

The framework suggested contains a five step process including (1) construction of patient vectors, (2) develop the semantic corpus, (3)build and train a neural network, (4) construct similarity profiles and (5) make predictions.

1. Innovation:   
   The innovation of this research is to enhance existing CDS systems. Some of the important lacking functionality in existing systems in which this paper was set out to improve include (1) the current solution focuses on a single patient at time while ignoring other similar patients (2) the decision to produce a diagnosis today is either manual or employs an automatic set of decision rules and (3) only a single medical condition is considered at a time.
2. Disadvantages/Advantages:   
   The advantage of this research was that it was trained on MIMIC data, however when the solution was performed on a real- world EHR dataset the results show that the approach is both effective and accurate. The solution also provides clinically meaningful interpretations.

The disadvantages of this research limits patient diagnosis to historically similar patients, there was little to no inclusion regarding diagnosis of unseen medical issues and how this may be dealt with.

1. Data Accessibility: Yes. MIMIC III v1.4, a publicly available critical care dataset. <http://staff.icar.cnr.it/diseaseDiagnosis.zip>
2. 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, JAYAKUMAR JEGANATHAN

1. 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.

1. Innovation:

The use of unstructured clinical notes and not just EHR data. A long-term aggregation system.

1. 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).

1. Data Accessibility: Yes. MIMIC III, a publicly available critical care dataset.
2. 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

1. 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.

1. 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.

1. Disadvantages/Advantages:

The model presents a high score overall categories. And after fine-tuning on a small sample the transfer model has a good performance.

1. Data Accessibility:

Yes, MIMIC-III data and UK-CRIS network data (the UK data access is dependent on receiving research approvals from NHS trust oversight bodies at Oxford Health NHS Foundation Trust )

1. Code Accessibility: Yes, https://github.com/kormilitzin/med7