



## Biomedical Informatics Special Seminar



*Imon Banerjee, PhD; Instructor, Department of Radiology and Biomedical Data Science, Stanford University*

### Natural Language Processing in Medicine: Annotation & Prediction

**Thursday, Apr 18, 2019 1 pm—2 pm**

**HSC-L3 Classroom 155**

#### Abstract:

The lack of labeled data creates a data “bottleneck” for developing deep learning models for medical imaging. However, healthcare institutions have millions of imaging studies which are associated with unstructured free text radiology reports that describe imaging features and diagnoses, but there are no reliable methods for leveraging these reports to create structured labels for training deep learning models. Unstructured free text thwarts machine understanding due to the ambiguity and variations in language among radiologists and healthcare organizations. Thus, most existing artificial intelligence (AI) applications in medicine, starting from diagnosis to prediction of treatment outcomes, largely exclude free-text narratives which are one of the richest component of electronic medical records (EMR). My talk will first describe recent developments of semantic word embedding methods to leverage narrative reports associated with radiological scans to generate automatically rich labels for training deep learning models and will highlight its application to difference domains of radiology (CT, MR, US, mammograms). I will also present recent AI projects that push beyond image classification and tackle the challenging problem of clinical outcome (e.g. progression and survival) prediction with longitudinal visit data, including images and free-text notes. The talk will briefly cover techniques for “peering inside the deep learning black box” to better understand what a deep learning model has learned for the task of making predications from longitudinal patient data. This direction of using longitudinal unstructured visit data will open a novel pathway for developing many advanced AI-based decision support systems that will leverage the unstructured information from medical “Big Data” to assist patient management.

#### Bio:

Imon Banerjee is an Instructor (junior faculty) in Department of Radiology and Biomedical Data Science at Stanford University. She was previously a research scientist in Biomedical Data Science department and completed her post-doctoral training from Department of Radiology at Stanford University. Her doctoral work in computer science focused on semantic annotation of medical data, and she received the Marie Curie European fellowship award and the Intel fellowship award for research on distributed computing in the European Organization for Nuclear Research. Her core expertise are unstructured data analysis and knowledge management and machine representation of image semantics. Her peer-reviewed publications relate to automated reasoning on unstructured and structured medical data. She is also creating a seamless fusion between high dimensional data analyses and semantic modeling techniques to generate patient-specific computational models for supporting treatment planning.

**Questions? Please call the Biomedical Informatics Department at 631-638-2590.**