## NIH/BD2K ISMB 2018:Machine Learning Approaches to Enable Biomedical Discoveries

## Automated Cohort Retrieval from EEG Medical Records

## Joseph Picone, Professor, Department of Electrical and Computer Engineering, Temple University

## Iyad Obeid, Professor, Department of Electrical and Computer Engineering, Temple University

## Sanda Harabagiu, Professor, Department of Computer Science, The University of Texas at Dallas

### Cohort retrieval systems can be a powerful clinical tool to aid diagnosis and training if we can harness the untapped potential of electronic medical records that include unstructured text, temporally constrained measurements (e.g., vital signs), multichannel signal data (e.g., EEGs), and image data (e.g., MRIs). In this talk, we will demonstrate a system that automatically ingests and organizes medical reports so that unstructured queries can be answered. Clinicians are able to retrieve relevant EEG signals and EEG reports using standard queries (e.g. “Young patients with focal cerebral dysfunction who were treated with Topamax”) from a large open source repository of clinical EEG data known as the TUH EEG Corpus. The system automatically annotates EEG signal data for events such as seizures and overall assessments such as abnormality using a multi-stage deep learning approach that integrates temporal and spatial context. Key clinical concepts are extracted from unstructured EEG reports using a novel recurrent neural network approach and used to understand semantic composition. The system also infers underspecified information and normalizes information across reports. Signal data annotations are combined with clinical concepts in a single unified representation based on a Qualified Medical Knowledge Graph. A novel learning-to-rank framework was developed to improve cohort ranking and usability based on relevance judgements produced by neurologists. The annotated data and knowledge representations are available as open source resources.