**Automatic Discovery and Processing of EEG Cohorts from Clinical Records**

Joseph Picone1, PhD; Iyad Obeid1; Sanda M. Harabagiu, PhD2

1 Neural Engineering Data Consortium, College of Engineering, Temple University, Philadelphia, Pennsylvania, U.S.A.

2 Human Language Technology Research Institute, University of Texas at Dallas, Dallas, Texas, U.S.A.

Decision support systems in healthcare can leverage vast archives of electronic medical records if high performance automated data wrangling can be achieved. EMRs can include unstructured text, temporally constrained measurements (e.g., vital signs), multichannel signal data (e.g., EEGs), and image data (e.g., MRIs). Our focus is the automatic interpretation of a clinical EEG Big Data resource known as the TUH EEG Corpus (TUH EEG). There are four major aims in this project: (1) automatically recognize and time-align events in EEG signals, (2) automatically recognize critical concepts in the EEG reports, (3) automatic patient cohort retrieval, and (4) evaluation and analysis of the results of the patient cohort retrieval.

We have developed two demonstrations of our cohort retrieval technology: a Multi-Modal EEG Patient Cohort Retrieval system called MERCuRY an acronym for Multi-modal EncephalogRam patient Cohort discoverY) and an EEG signal visualization tool. We have developed novel methods of identifying in the EEG reports the EEG activities, EEG events and patterns as well as their attributes. In addition to the EEG-specific medical concepts, we have also identified through our methods all medical concepts that describe the clinical picture and therapy of the patients. We have validated the usefulness of the patient cohort identification system by collecting feedback from clinicians and medical students.

Identification of the type and temporal location of EEG signal events such as spikes or generalized periodic epileptiform discharges in the EEG signal are critical to the interpretation of an EEG. We have developed a high performance event detection system that integrates hidden Markov models for event detection and deep learning for postprocessing. Our performance on a standard clinical event recognition task has improved to 91.4% sensitivity with an 8.5% specificity, which is approaching the level of performance required by clinicians.

Submitting Author’s career stage: Professor