**EEG Cohort Retrieval**

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. Electronic medical records (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). We have developed a demonstration of cohort retrieval technology that supports free text queries of a database of EEG signals and medical records. Users can construct open-ended queries (e.g., “young patients with focal cerebral dysfunction who were treated with Topamax”) or use a simple query-building interface. The system returns a ranked list of EEG sessions that best match the search criteria. There are two main components to the system: (1) AutoEEG: a deep learning-based system that identifies a variety of signal events, some of which are global judgements (e.g., abnormality) and some of which are local events (e.g., a seizure or spike); and (2) MERCuRY: a Multi-Modal EEG Patient Cohort Retrieval system that integrates domain knowledge as well as temporal and spatial relationships. We have developed novel methods of identifying EEG events and patterns as well as their attributes. In addition to the EEG-specific medical concepts, we have also identified 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.