**Temple University:**

The Temple University team employed a large number of students and one postdoc on this project. There were four major labor categories:

1. Postdoc: Scott Yang was the sole postdoc recruited for this project. Since he joined the project with limited programming experience (primarily MATLAB), he received intensive training on a variety of software skills: Unix/Linux command line programming, C/C++ programming, Python programming and html/web programming. He was also introduced to a rigorous software engineering process including revision control software and GitHub/GitLab. Further, he received training on state-of-the-art machine learning packages such as Theano, Kera, TensorFlow and HTK.

With respect to professional development, in addition to learning how to a set of advanced features in Microsoft Office and Mendeley to develop and manage publications, he received significant training on proposal writing, technical presentations and technical writing.

He was encouraged to publish and attend professional conferences. These are described in the extensive publication list presented later in this report.

1. Graduate Students: We have employed six graduate students on this project. Four were pursuing PhDs and two pursued MS degrees. Therefore, an important part of their training included normal graduate student training such as didactic graduate-level courses, thesis and dissertation proposals and PhD preliminary exams.

As with (1), they also received significant amounts of programming training and professional development training through publications and conference presentations. Further, they led the development of several real-time demonstration systems and were trained how to develop and present technical demonstrations.

Training was so successful that we lost two of these students to industry jobs before they graduated. Their marketability was a testament to their training, as students trained in machine learning are received very attractive job offers from industry these days. Both students will complete their PhDs remotely, but working full-time certainly slows the process down.

1. Undergraduate Students: We employ a large number of undergraduates on this project in four capacities: (1) data annotation, (2) software engineering, (3) web development and (4) IT support. The data annotators learn how to annotate EEGs in a manner similar to the way neurologists annotate the data. We have a well-developed process for training them and in a publication cited later we show that their annotation accuracy is excellent compared to clinically-trained neurologists.

The software engineering training is similar to what was previously described. Our undergraduates become expert programmers in C/C++ and Python, which makes them very valuable. We assist them in obtaining summer internships and full-time employment once they are trained. It is very clear that they qualify for jobs they would not normally be able to get after this training.

Several of the students work in positions we described as IT support, which involves maintaining our Linux cluster and server environment, and web support, which is the primary means by which we disseminate information and manage customer relationships. Needless to say, these skills, which typical ECE students at Temple don’t develop, make them extremely marketable in sectors ranging from defense to financial engineering.

Finally, all students involved in this project received extensive training on how to read and interpret EEGs. They worked closely with the Department of Neurology at Temple Hospital and attended many EEG readings sessions conducted by the hospital as part of their medical student training.

**University of Texas at Dallas:**

Six PhD students have been advised for their research conducted for this project at University of Texas at Dallas.

Travis Goodwin has defended his PhD thesis in March 2018, after 6 years of PhD studies in Computer Science at UTD, advised by Prof. Sanda Harabagiu. He has accepted a fellowship at NIH starting in June 2018. He participated in our project for the past three years, developing novel research in the area of multi-modal indexing, inference of underspecified information in the EEG reports and interaction of various factors in the EEG reports. In November 2017, he received the **Homer Warner award** at the 2017 AMIA Symposium for the paper co-authored with his adviser, Dr. Sanda Harabagiu, entitled “*Inferring Clinical Correlations from EEG Reports with Deep Neural Learning*”. Travis has also been working on defining the HAD tags under the supplement project. He has also worked on using deep learning methods for the automatic annotation of HAD tags as well for generating data-driven neural knowledge representations of the knowledge discerned from the EEG reports. He is also the co-author of a paper that received the **AMIA Clinical Research Informatics Award** at the 2018 AMIA Informatics Summit, in March 2018. The paper is entitled “*Memory-Augmented Active Deep Learning for Identifying Relations Between Distant Medical Concepts in Electroencephalography Reports”* with the authors: Ramon Maldonado, Travis Goodwin and Sanda Harabagiu. Travis has successfully submitted 23 conference papers and has received in 2016 the **Best Student Paper Award** at the ACM International Conference in Information and Knowledge Management (CIKM-2016), a major conference on knowledge managements and information retrieval. In addition, he has published four journal papers. These accomplishments exceeded Travis’s Individual Development Plans (IDPs).

Ramon Maldonado is a 3rd year PhD student in Computer Science at UTD, advised by Prof. Sanda Harabagiu, who has performed research on automatically identifying all medical concepts and recognizing the relations between them in the Temple University Hospital EEG data, in the form of EEG reports documenting 25,000 sessions and 15,000 patients collected over 12 years at Temple University Hospital. Ramon is a qualified PhD student, by passing a set of qualifying exams, while developing new techniques for his research, which is remarkable. He is the lead author on the paper entitled “*Memory-Augmented Active Deep Learning for Identifying Relations Between Distant Medical Concepts in Electroencephalography Reports”,* which has received the **AMIA Clinical Research Informatics Award** at the 2018 AMIA Informatics Summit, in March 2018. Ramon is also the lead authors of the paper entitled *“Active Deep Learning-Based Annotation of Electroencephalography Reports for Cohort Identification”,* which was **nominated for** **Distinguished Paper Award** at the 2017 American Medical Informatics Association Joint Summits on Clinical Research Informatics (AMIA-CRI*).*  Ramon has also been lead author on a paper published in *the American Medical Informatics Association Annual Symposium (AMIA-2017)* which introduced a novel form of medical knowledge, called *medical knowledge embeddings*, developed from the TUH corpus, in the form of EEG-MKE. The EEG-MKE have been described in the paper entitled “*Deep Learning Meets Biomedical Ontologies: Knowledge Embeddings for Epilepsy*”. Moreover, Ramon was a co-author on a journal paper published this year in the Journal of Biomedical Informatics. These accomplishments meet Ramon’s Individual Development Plans (IDPs).

Stuart Taylor is a 2nd year PhD student in Computer Science at UTD, advised by Prof. Sanda Harabagiu, who has worked on the initial development of the active deep learning for annotating EEG reports. Stuart is a qualified PhD student, by passing a set of qualifying exams, while developing new techniques for his research, which is remarkable. He worked on the generation of queries for the evaluation of patient cohorts. Stuart has published a poster at the *American Medical Informatics Association Annual Symposium (AMIA)* in 2017 and submitted a full paper at the *American Medical Informatics Association Annual Symposium (AMIA)* in 2018. has plan on working on the recognition of HAD tags in biomedical texts. These accomplishments meet Stuart’s Individual Development Plans (IDPs).

Pracheta Sahoo, Shasha Jin and SaraRouhani, are qualified PhD students in Computer Science at UTD, advised by Prof. Vibhav Gogate, who have participated in the project a few months this year to help with machine learning frameworks for deep learning and for evaluation of the results as well as to produce relevance judgements. They are part of the Statistical Relational Artificial Intelligence and Machine Learning Lab, led by Prof. Gogate at UTD. They have published papers in the AAAI conference.