In the real world, data used to build machine learning models always has different sizes and characteristics. These size and characteristic features, including small datasets, big datasets, imbalanced datasets, often lead to different challenges when training machine learning models. Models trained on a small number of observations tend to overfit the training data and produce inaccurate results. When it comes to big data, efficiently learning from "huge" size data in a short time becomes important. With an imbalanced dataset, learning is usually biased towards the majority class in the data and appropriate measurements are needed to check model performance. In medical imaging, we proposed a novel image feature extraction method for predicting survival time from brain tumor magnetic resonance images using pre-trained deep neural networks. We also introduced a novel method for over-sampling the minority class examples at the image level, rather than the feature vector level, to provide a solution to the problem of imbalanced medical imaging data. For social network analysis and future forecasting, we introduced a decomposition approach to address the long term fine time granularity simulation problem. The goal is to predict different user activities at hour granularity over a long period of time. In addition, when considering simulating user activities across multiple platforms, we introduced a sequence model approach which provides efficient long term cross platform simulation.

### Publications


### Disability Accommodations:

If you require a reasonable accommodation to participate, please contact the Office of Diversity & Equal Opportunity at 813-974-4373 at least five (5) working days prior to the event.