1131-62-286 **Jared Lee Ostmeyer*** (jared.ostmeyer@utsouthwestern.edu), 5323 Harry Hines Blvd, Dallas, TX 75390-9066, and **Scott Christley**, **William Rounds**, **Inimary Toby**, **Nancy Monson** and **Lindsay Cowell** (lindsay.cowell@utsouthwestern.edu), 5323 Harry Hines Blvd, Dallas, TX 75390-9066. *Machine Learning on sets and sequences and its applications in diagnosing disease.*

We will present our methods for performing statistical classification on labeled sets and labeled sequences. Our overall approach is to score each item in a set or each symbol in a sequence using a parameterized scoring function and to aggregate the scores into a predicted label. When the items are permutation invariant, as is the case with a set, a generalized mathematical mean is used to reduce the scores to a single value and predict a label. When dealing with symbols in a sequence, the ordering of the symbols is critical information, and so we introduce the idea of aggregating the scores from each symbol using a recurrent weighted average. In either case, once a predicted label is obtained for each data point, we can define the likelihood function and use standard optimization techniques to determine specific values for the model's parameters. As a practical application of our research, we will show how to build a simple statistical classifier that takes a set of immune receptor sequences as input and reduces the data to a predicted diagnosis of either Multiple Sclerosis or other neurological disease. On unseen test data collected separately from our training data, our model achieves $\sim$75% accuracy, an improvement over the current standard diagnostic approach. (Received July 17, 2017)