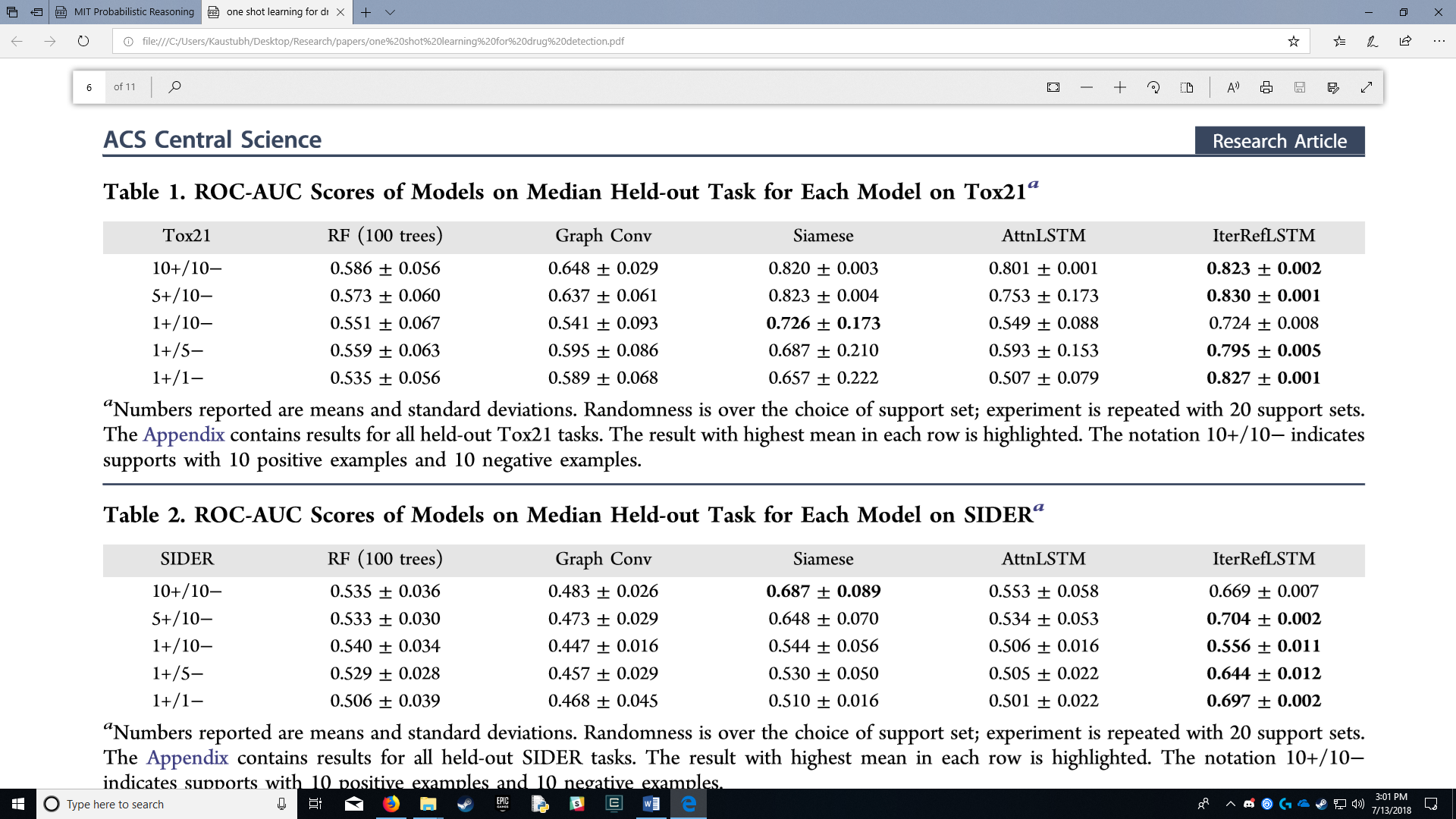
Notes for small data

Cant figure out the math on this one, but seems to be very promising. Part of a new architecture for one shot deep learning/probabilistic models. (Iterative Reﬁnement LSTMs)



Problems:

1. Outliers/Noise
2. Overfitting

Ideas:

**Do use Regularization**

Regularization is an almost-magical solution that constraints model fitting and reduces the effective degrees of freedom without reducing the actual number of parameters in the model. **L1 regularization** produces models with fewer non-zero parameters, effectively performing implicit feature selection, which could be desirable for explainability of performance in production, while **L2 regularization** produces models with more conservative (closer to zero) parameters and is effectively similar to having strong zero-centered priors for the parameters (in the Bayesian world). L2 is usually better for prediction accuracy than L1.