UNIVERSITY OF CAMBRIDGE

# Active Learning with High Dimensional Inputs using Bayesian Convolutional Neural Networks

Riashat Islam

Supervisors: Zoubin Ghahramani and Yarin Gal

# Outline of Experiments

- Performance using Dropout Acquisition Functions when used with Bayesian ConvNet and avoid over-fitting when using small datasets
- Comparison of Dropout Acquisition Functions with Other Baseline active learning acquisition functions
- Significance of Query Rate per acquisition
- Comparison of performance using only softmax and uncertainty estimates from test-time dropout
- Significance of model architecture and non-linearity for different uncertainty estimates for use of active learning with deep models
- Comparison with combination of SSL and AL (Minimum Bayes Risk for Binary Classification)
- Comparison of our approach with other recent methods for data-efficiency in deep learning
- Acquisition Functions using Dropout + Random uncertainty estimates in active learning

# Acquisition Functions

**Dropout Acquisition Functions**

- Dropout Bayesian Active Learning By Disagreement
- Dropout Variation Ratio
- Dropout Bayes Segnet
- Dropout Maximum Entropy
- Dropout Least Confident

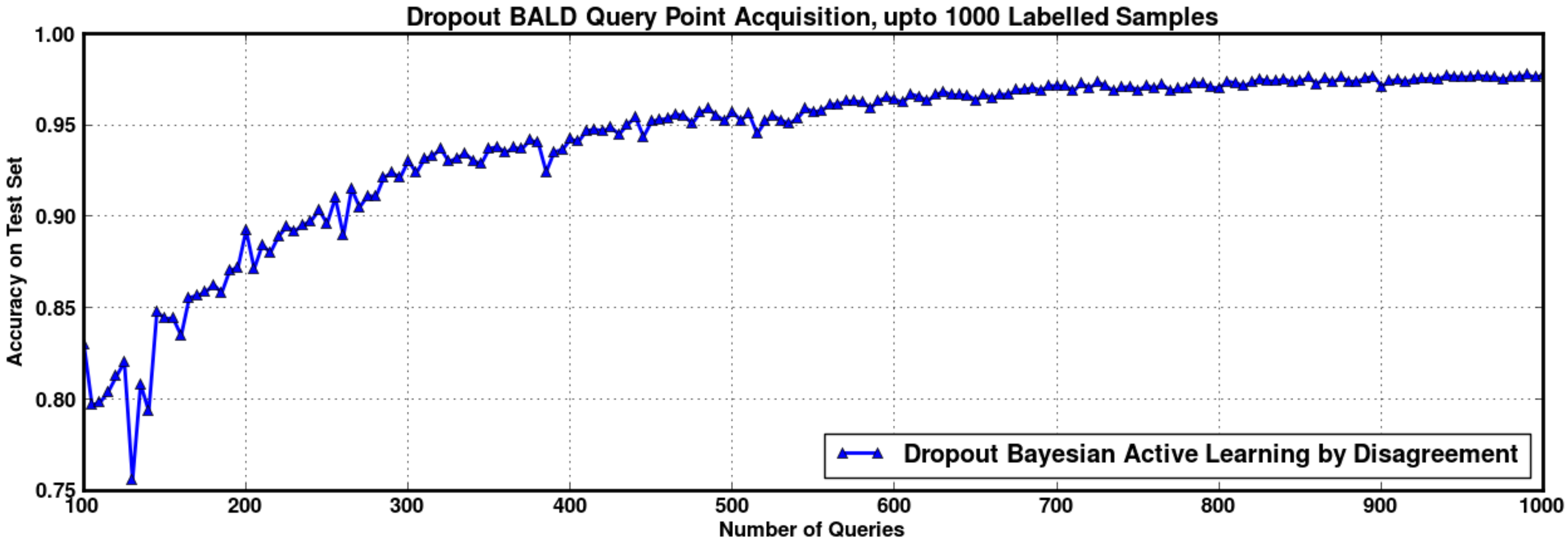
**Other baseline acquisition functions used in active learning**

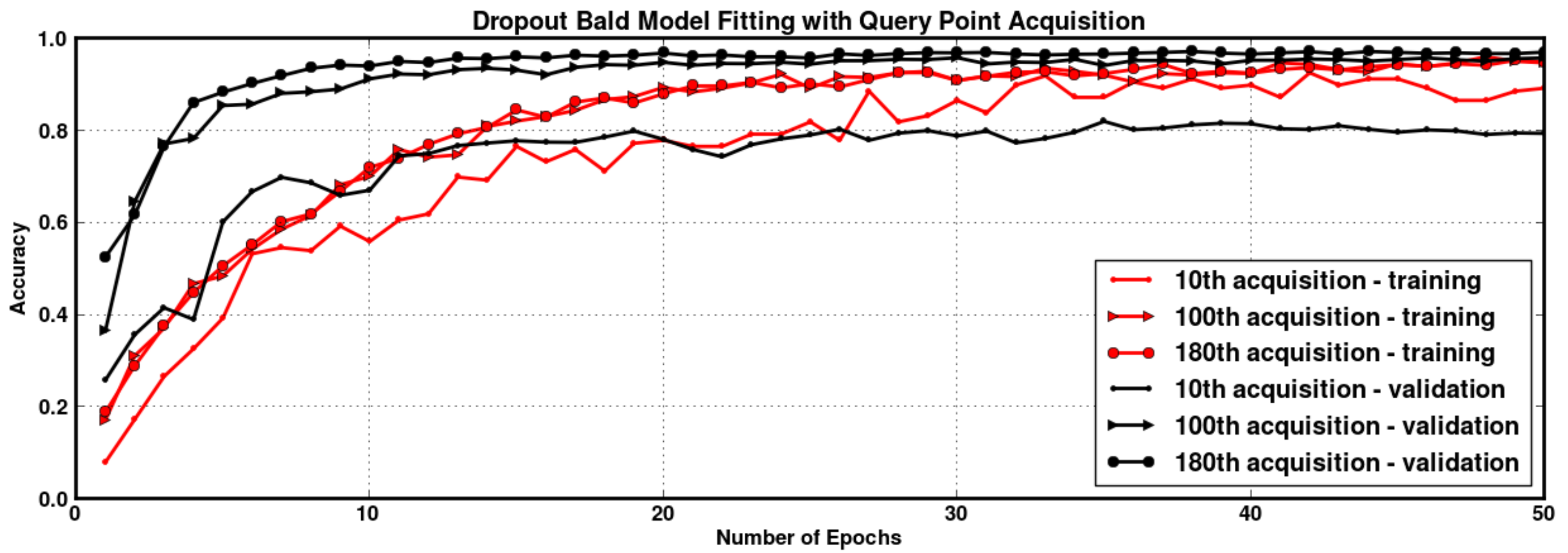
- Uncertainty Sampling – Maximum Entropy
- Uncertainty Sampling – Best vs Second Best Search (BvSB)
- Maximum Entropy
- Random acquisition

# Experimental Setup

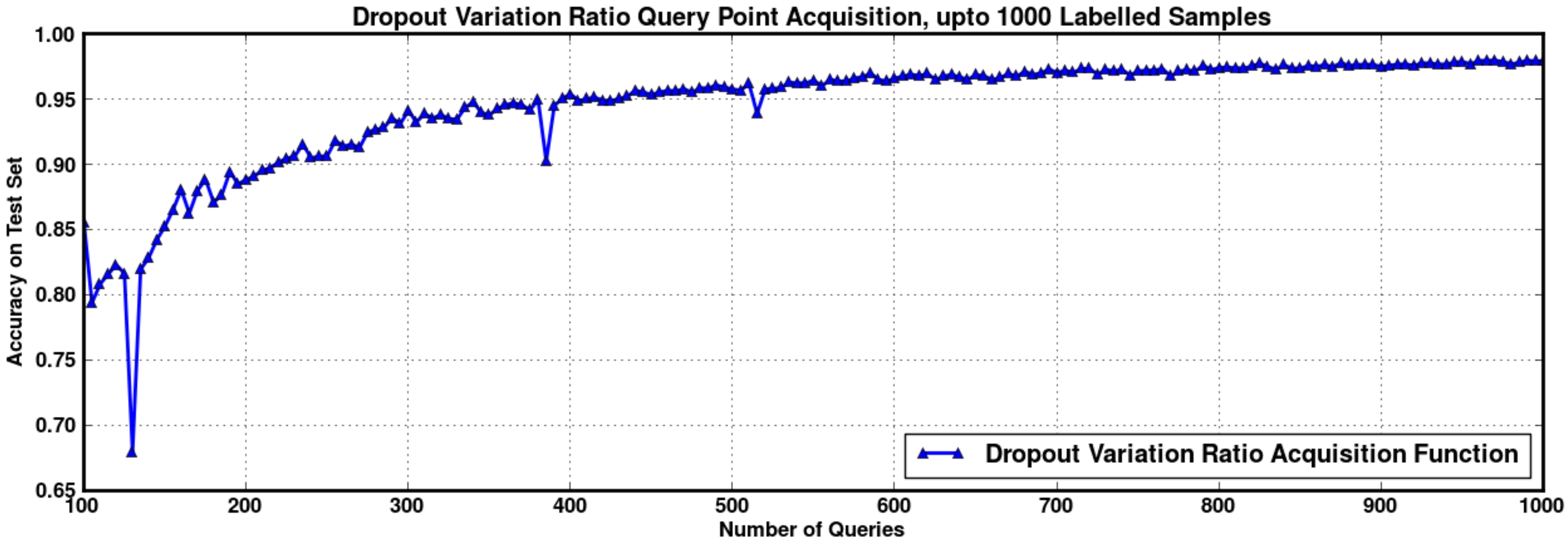
- Number of Experiments = 3-5 for averaged results
- LeNet5 architecture
- Number of Epochs = 50
- Starting training data: 20 - 100 data points, using up to 1000 training labelled samples
- 10,000 test samples on MNIST
- Number of Queries made at each acquisition: 1, 5 or 10
- 100 Dropout MC Samples for uncertainty estimates
- Weighted inputs in the loss function
- ADAM optimizer

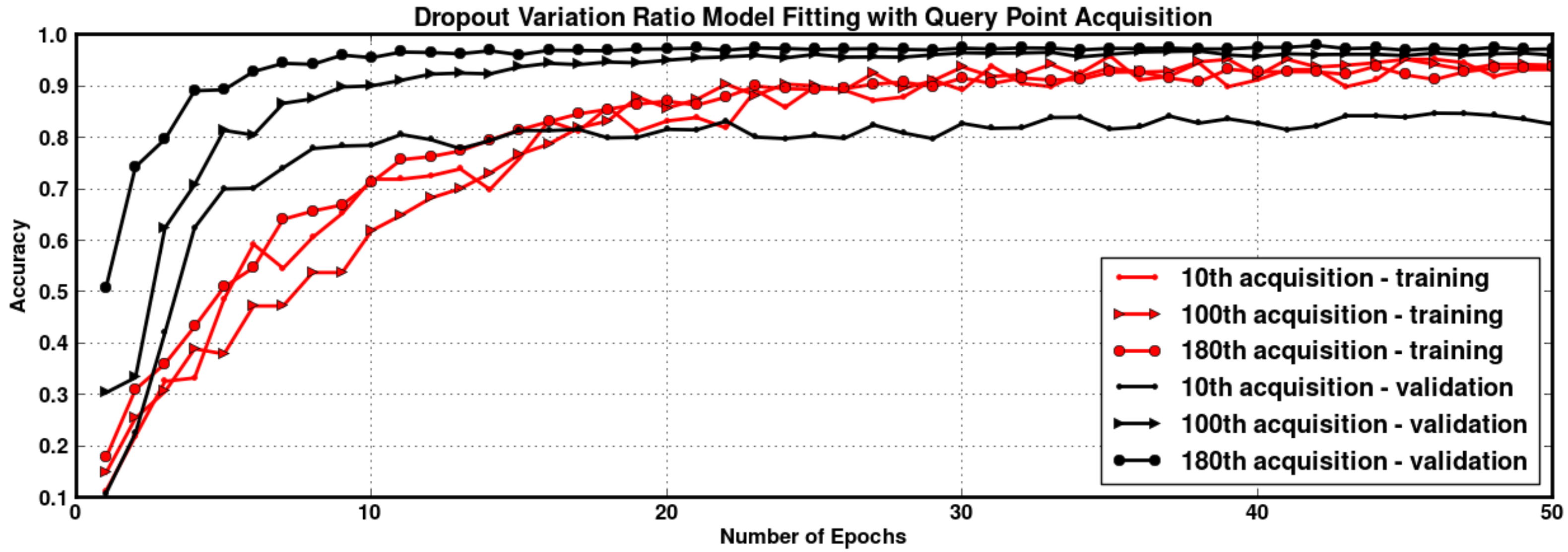
# Dropout Bayesian Active Learning by Disagreement



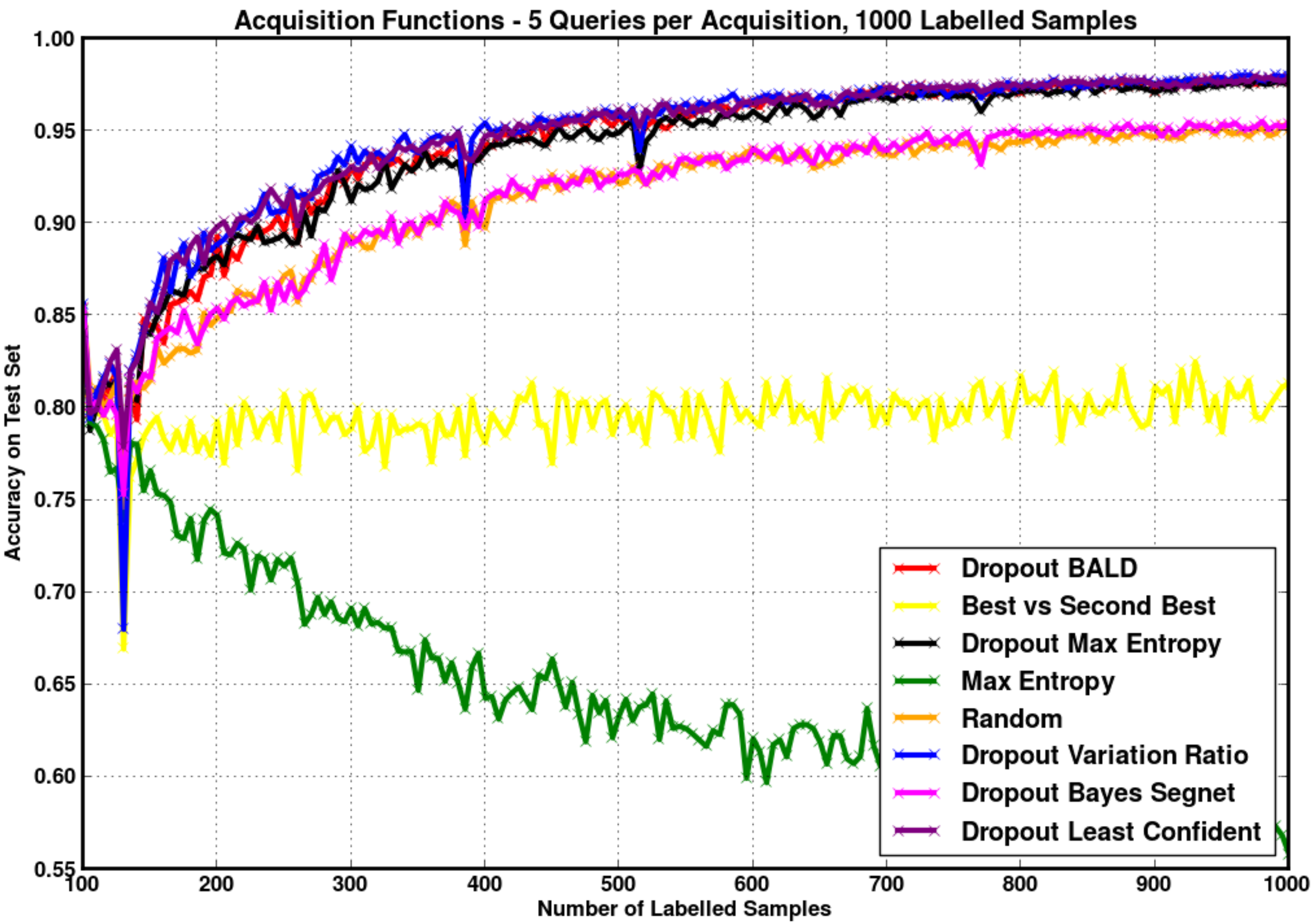



# Dropout Variation Ratio

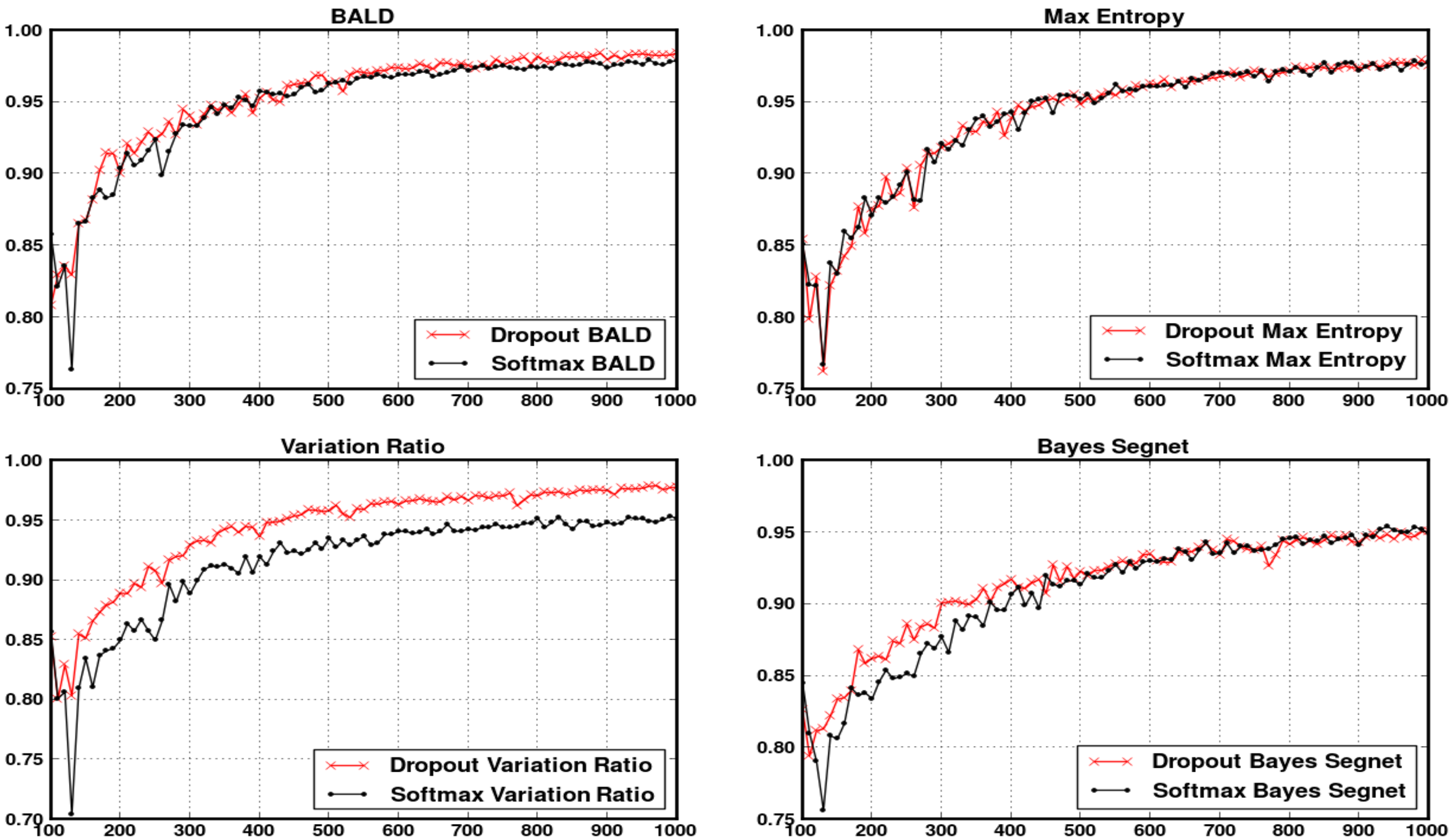


# Comparison of Acquisition Functions

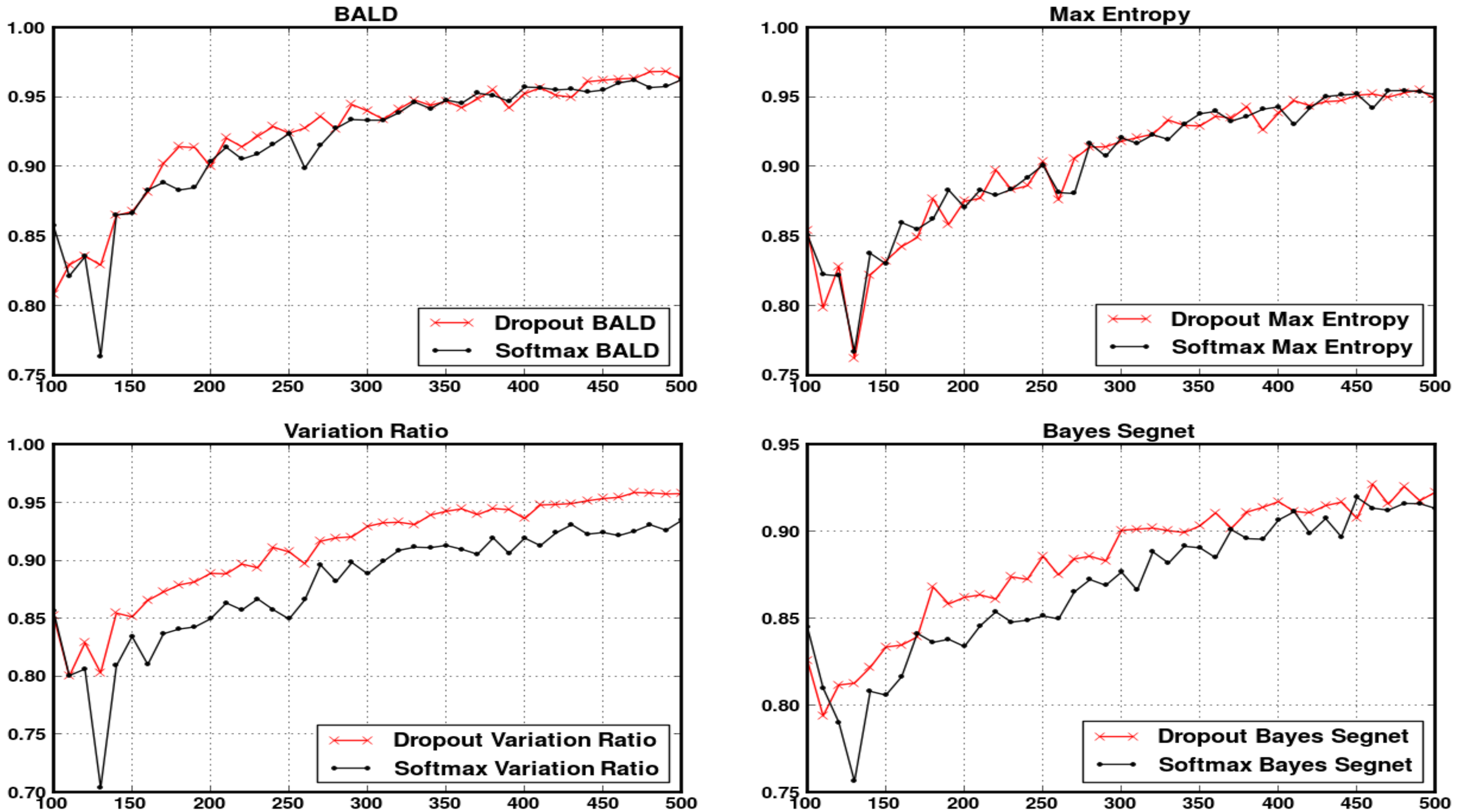


# Significance of Uncertainty Estimates

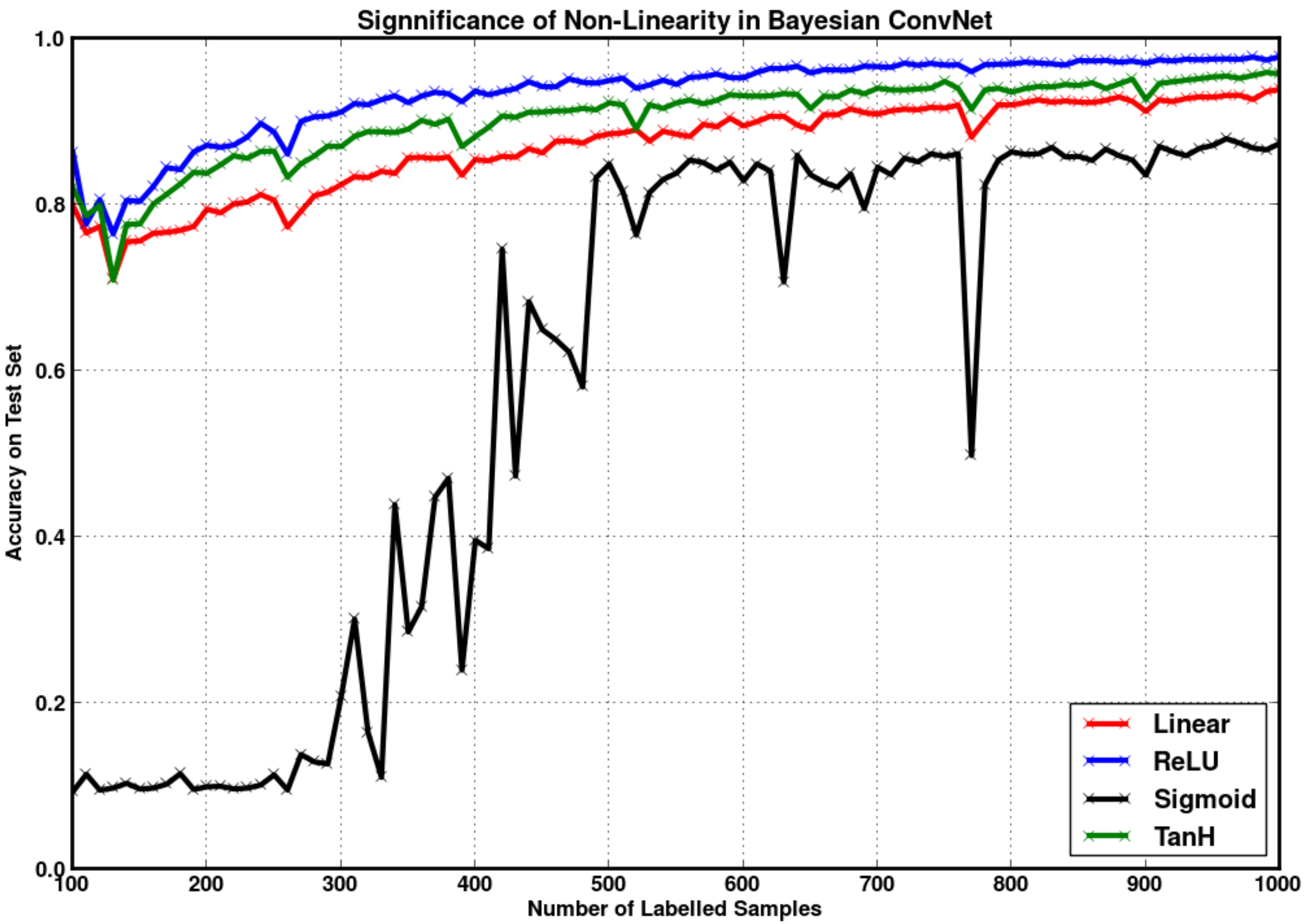


# Significance of Uncertainty Estimates

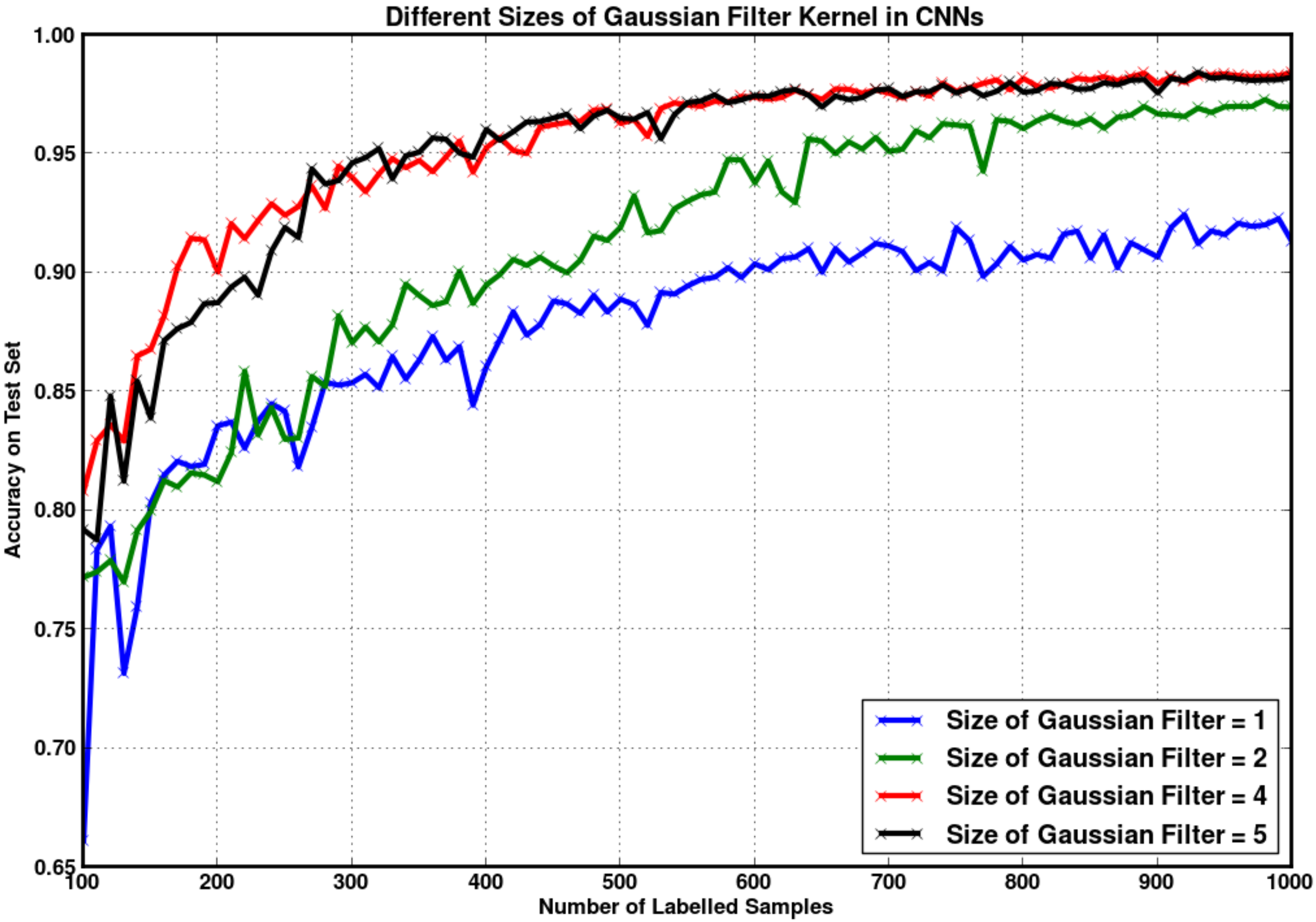
## Effect of test-time dropout more significant in small data settings



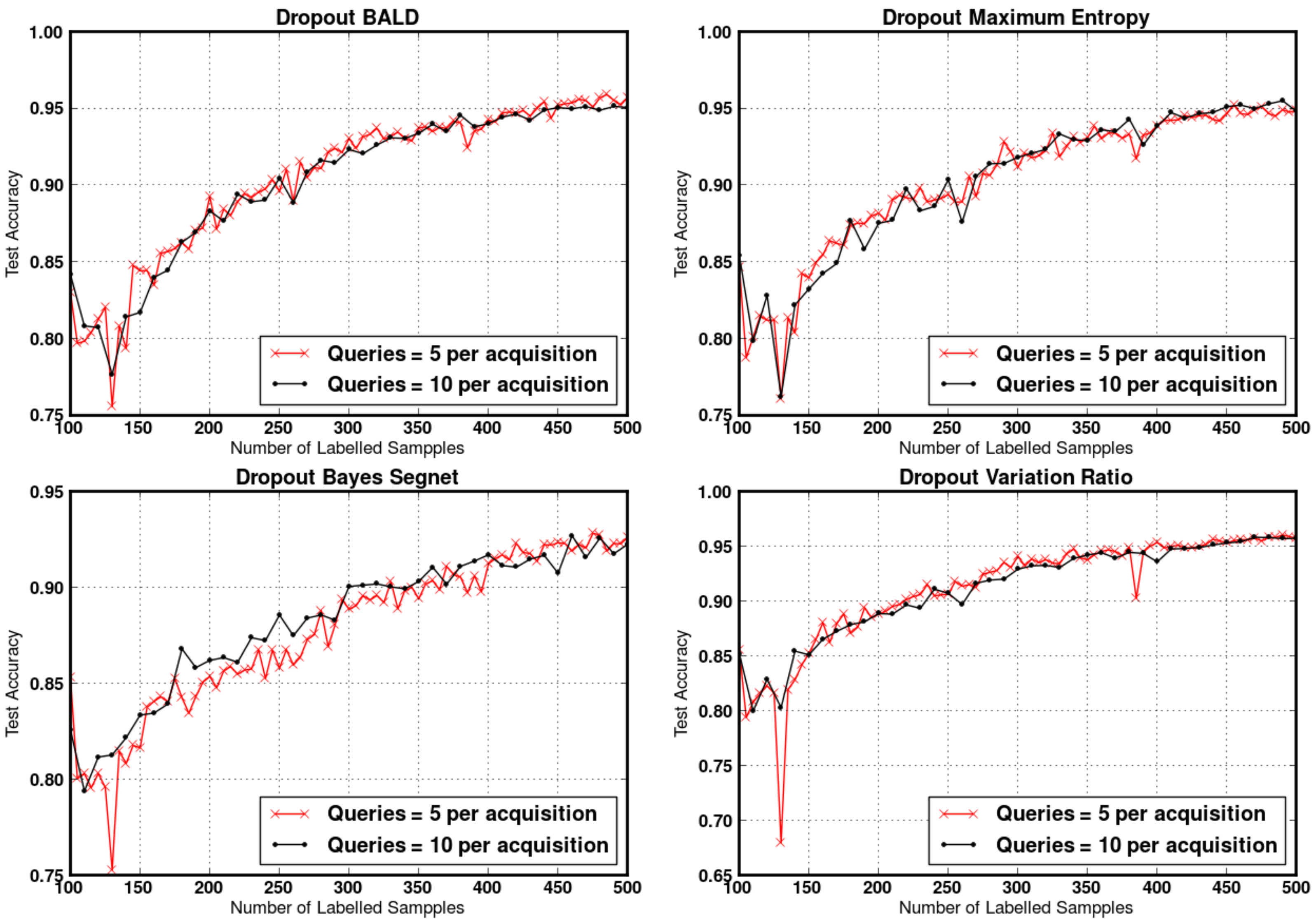
# Non-Linearity



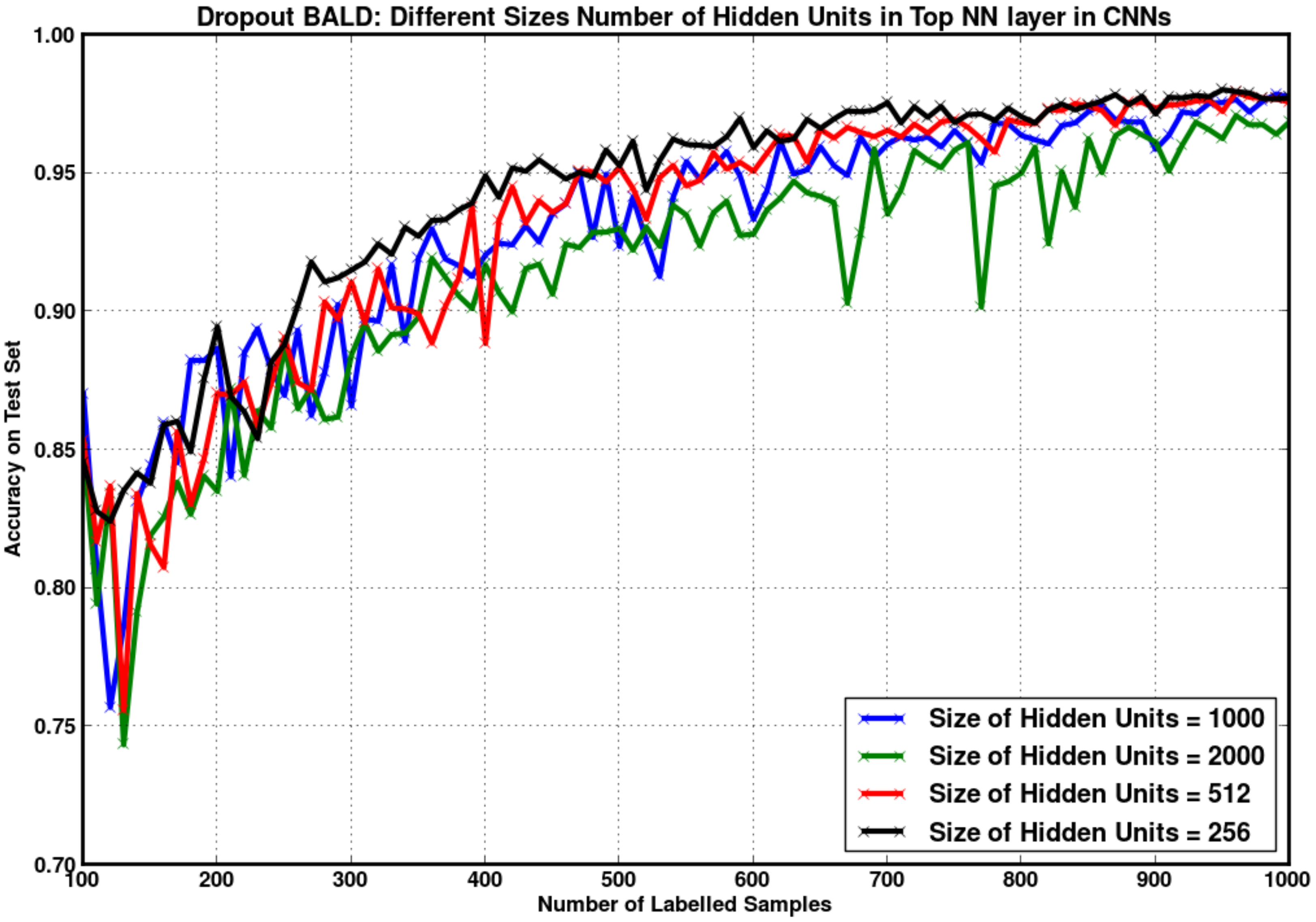
# Kernel Filter Size



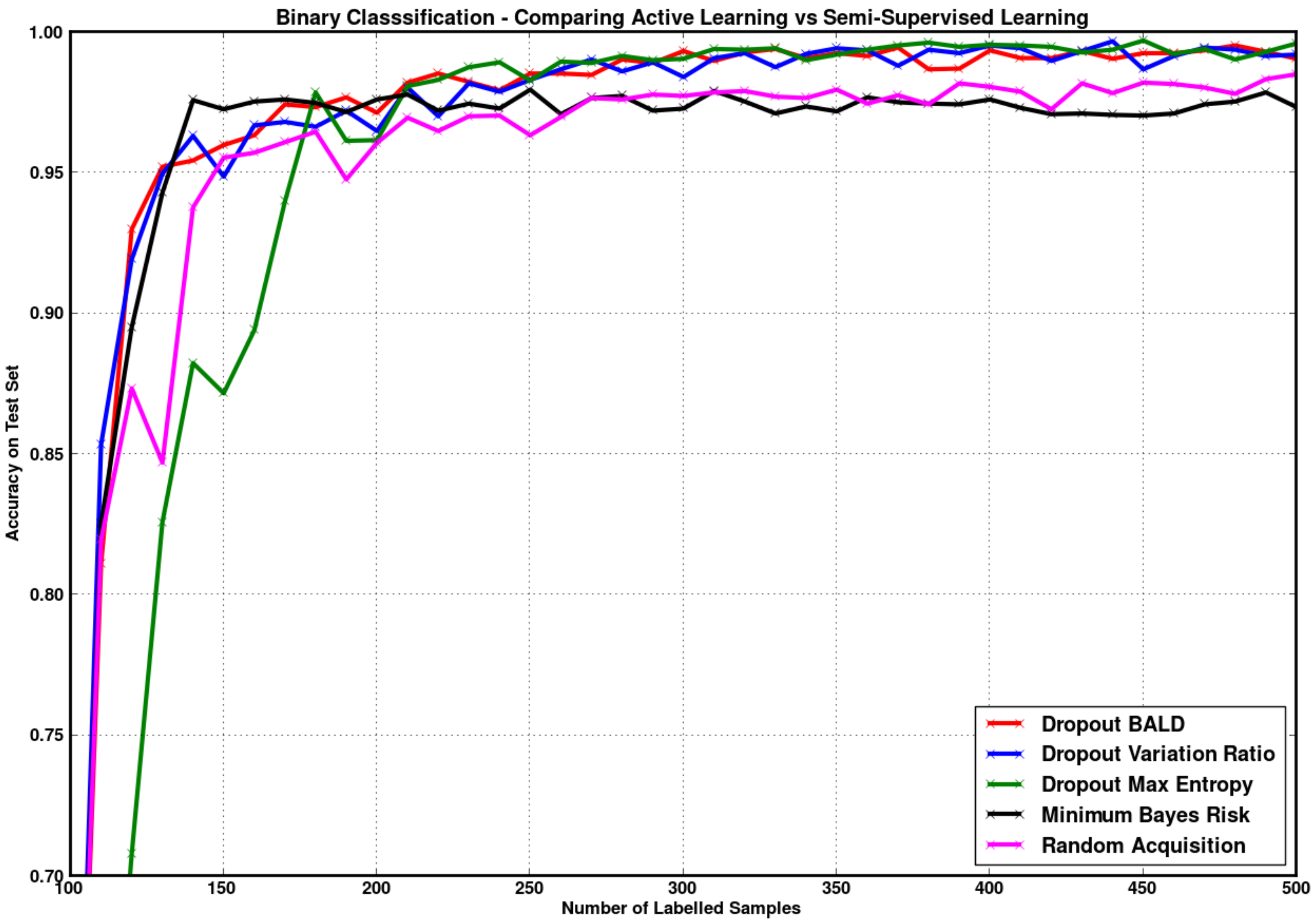
# Significance of Query Rate for Computational Efficiency



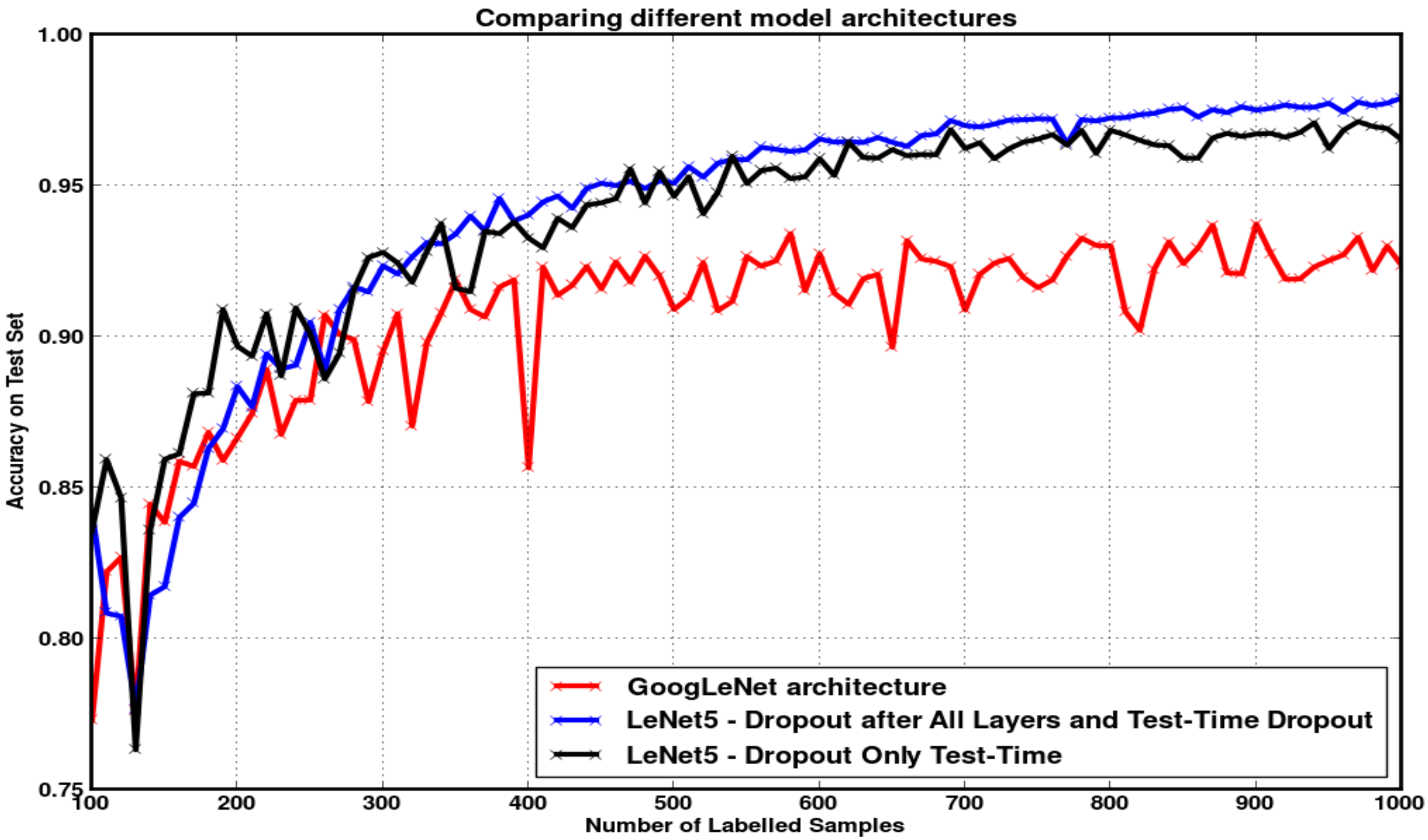
# Number of Hidden Units



# Binary Classification – Comparing with Minimum Bayes Risk



# Model Architectures



# Comparison with SSL Methods

**Test Error Results on MNIST for 1000 labelled training samples**

| Test error % on 10,000 samples with number of used training labels | 1000 |
|---|---|
| Semi-sup. Embedding (Weston et al., 2012) | 5.73 |
| MTC (Rifai et al., 2011) | 3.64 |
| Pseudo-label (Lee, 2013) | 3.46 |
| AtlasRBF (Pitelis et al.,2014) | 3.68 |
| DGN (Kingma et al., 2014) | 2.40 |
| Virtual Adversarial (Miyato et al., 2015) | **1.32** |
| SSL with Ladder Networks (Rasmus et al., 2015) | **0.84** |
| Dropout BALD | **1.57** |
| Dropout Variation Ratio | **2.05** |
| Dropout Maximum Entropy | **2.37** |
| Dropout Least Confident | **2.14** |
| Dropout Bayes Segnet | **4.62** |

# Summary of Results

**Test Error Results on MNIST for 100, 1000 and 3000 labelled training samples**

| Test accuracy % on 10,000 test samples with number of used training labels | 100 | 1000 | 3000 |
|---|---|---|---|
| Dropout BALD | - | 98.43 | 98.84 |
| Dropout Variation Ratio | - | 97.95 | 98.87 |
| Dropout Maximum Entropy | - | 97.63 | 98.84 |
| Dropout Least Confident | - | 97.86 | 98.87 |
| Dropout Bayes Segnet | - | 95.38 | 97.19 |
| Random Acquisition | - | 94.95 | 97.31 |
| Uncertainty Sampling (Max Margin aka BvSB) | - | 83.95 | 82.77 |
| Uncertainty Sampling (Max Entropy) | - | 53.28 | 36.10 |

# Dropout + Random Uncertainty Estimates