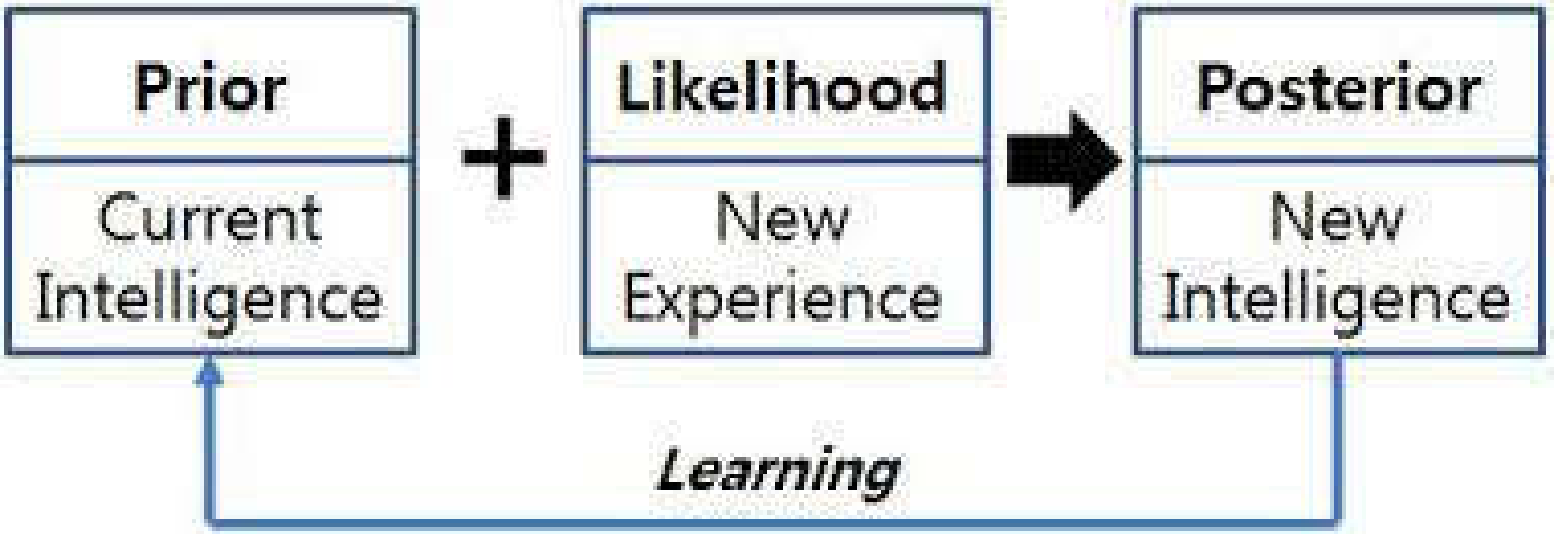
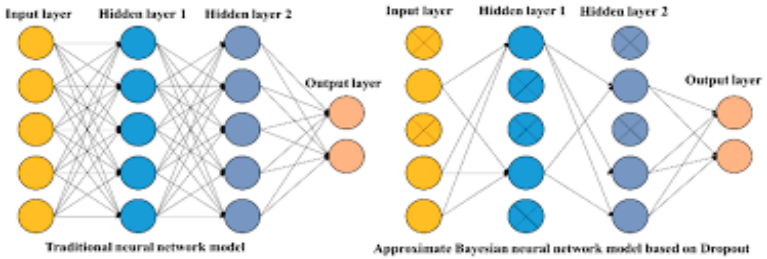
Bayesian Artificial Learning

Babacar Diatta and Jacob Olinger





Traditional Deep Learning

- Struggles with Difficult-to-Label data sets
- Scalability Issue

Bayesian Active Learning

Why not take the data that will have the most impact?

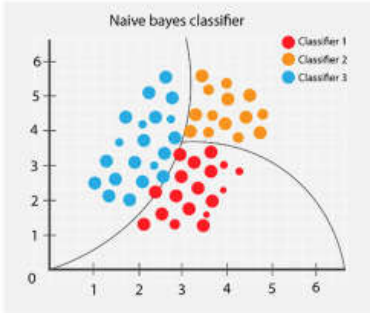
Naive Bayes

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$



Bayesian Active Learning

- Entropy Vs. Bald Key Differences:

- Entropy only considers the uncertainty in the model's predictions, while BALD considers the disagreement between the entire posterior belief and prior over models.
- So BALD considers how much the model's belief distribution p(ω|D) over all possible models/hypotheses ω has shifted away from the initial prior p(ω) after seeing the data D.
- For complex model classes like neural nets, the BALD criterion has been shown to be more data-efficient than entropy sampling.
- However, estimating BALD scores can be computationally more expensive than entropy for approximations.

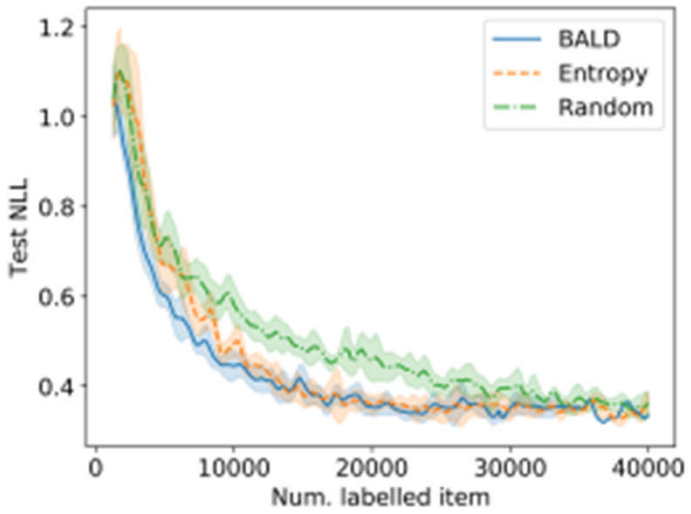
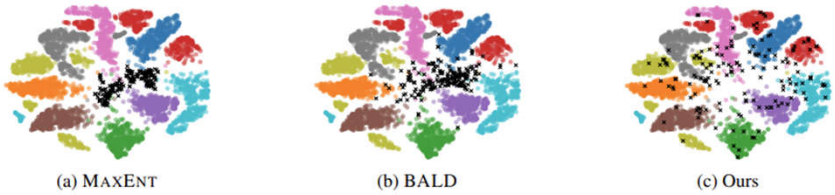


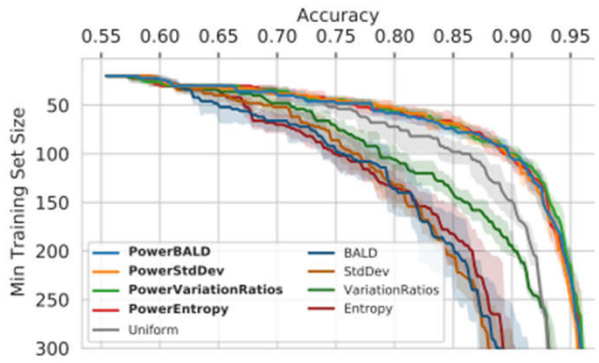
Figure 1. Baselines results on CIFAR10 using MC-Dropout and VGG-16. On an academic dataset, both active learning techniques are competitive.

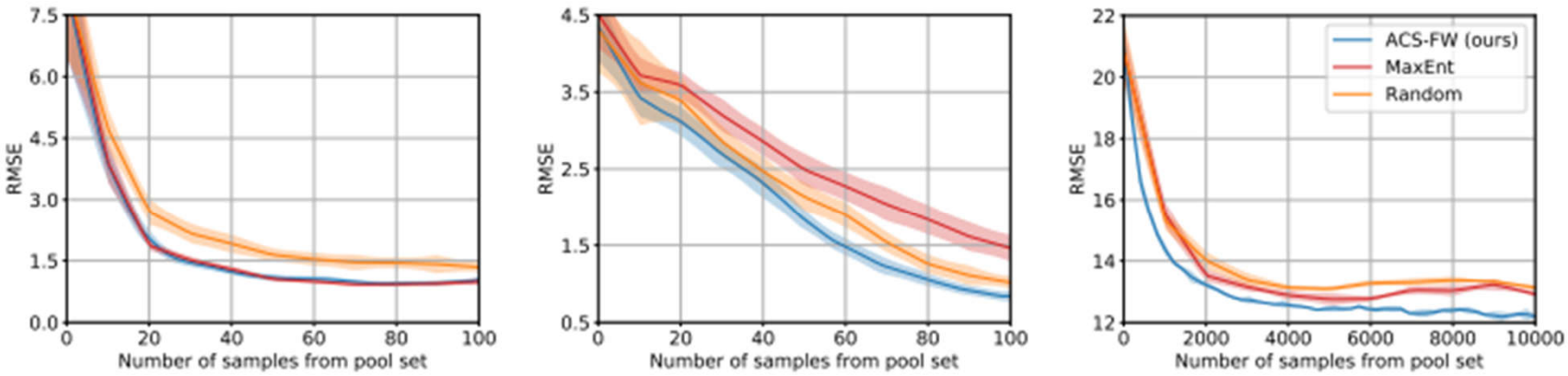
Batch Processing



(a) MaxEnt (b) BALD (c) Ours

- Traditional Bayesian Active Learning relies on selecting data points one at a time.
- This sequential process can be slow and computationally demanding.
- Advanced batch-processing techniques have been developed to improve efficiency.
- Batch processing involves selecting groups of data points simultaneously.
- By doing so, batch processing accelerates the learning process.
- It also enhances the quality of learning by reducing the impact of outliers.
- Additionally, batch processing minimizes redundancy in selected samples.





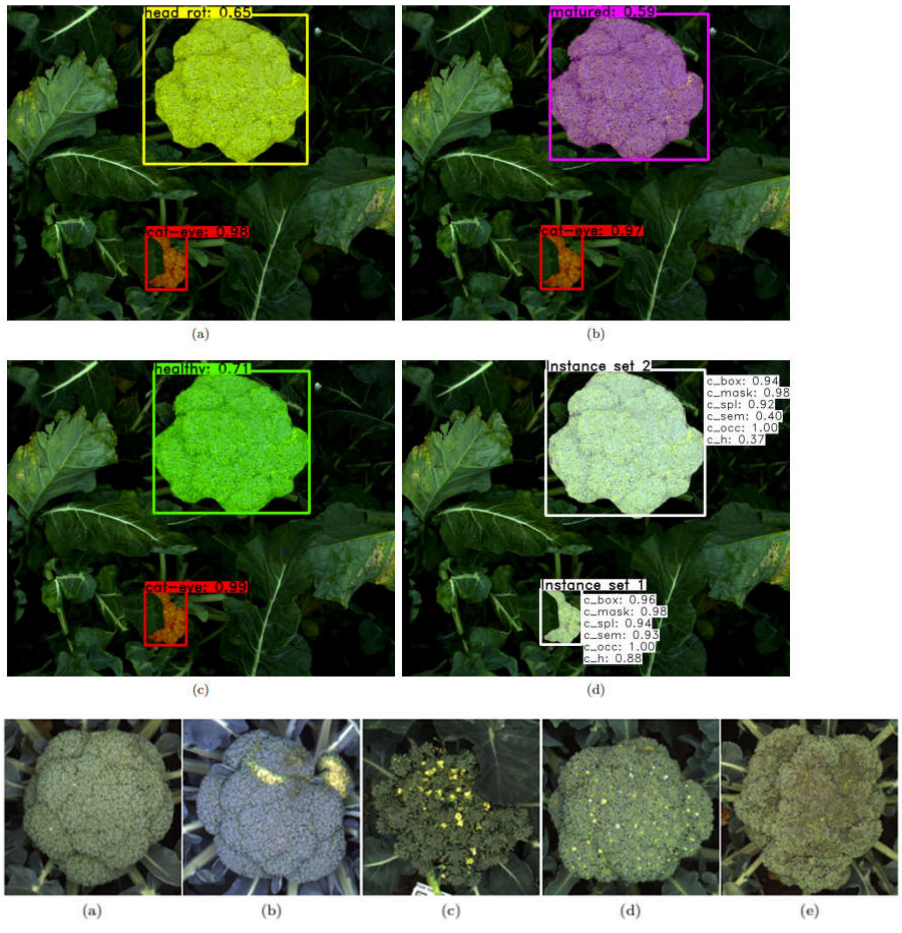
(a) yacht (b) energy (c) year

Sparse Set Approximation

- Sparse Subset Approximation is an advanced technique in batch processing.
- It aims to approximate the log posterior of the entire dataset using a smaller subset of data.
- This approach reduces computational demands by focusing on key data points.
- The Frank-Wolfe optimization algorithm is utilized for efficient batch construction.
- It's particularly effective for large-scale datasets and high-dimensional spaces.

Stochastic Methods

- Stochastic methods introduce a probabilistic element into batch selection.
- These methods enhance traditional random sampling by weighting selection probability.
- The selection probability is based on the posterior probability of each data point.
- This approach balances exploration and exploitation effectively.
- It prevents the model from fixating on familiar regions of the data space.
- Stochastic methods are beneficial for managing complex data structures.
- They promote diverse and representative batch selections for robust training.

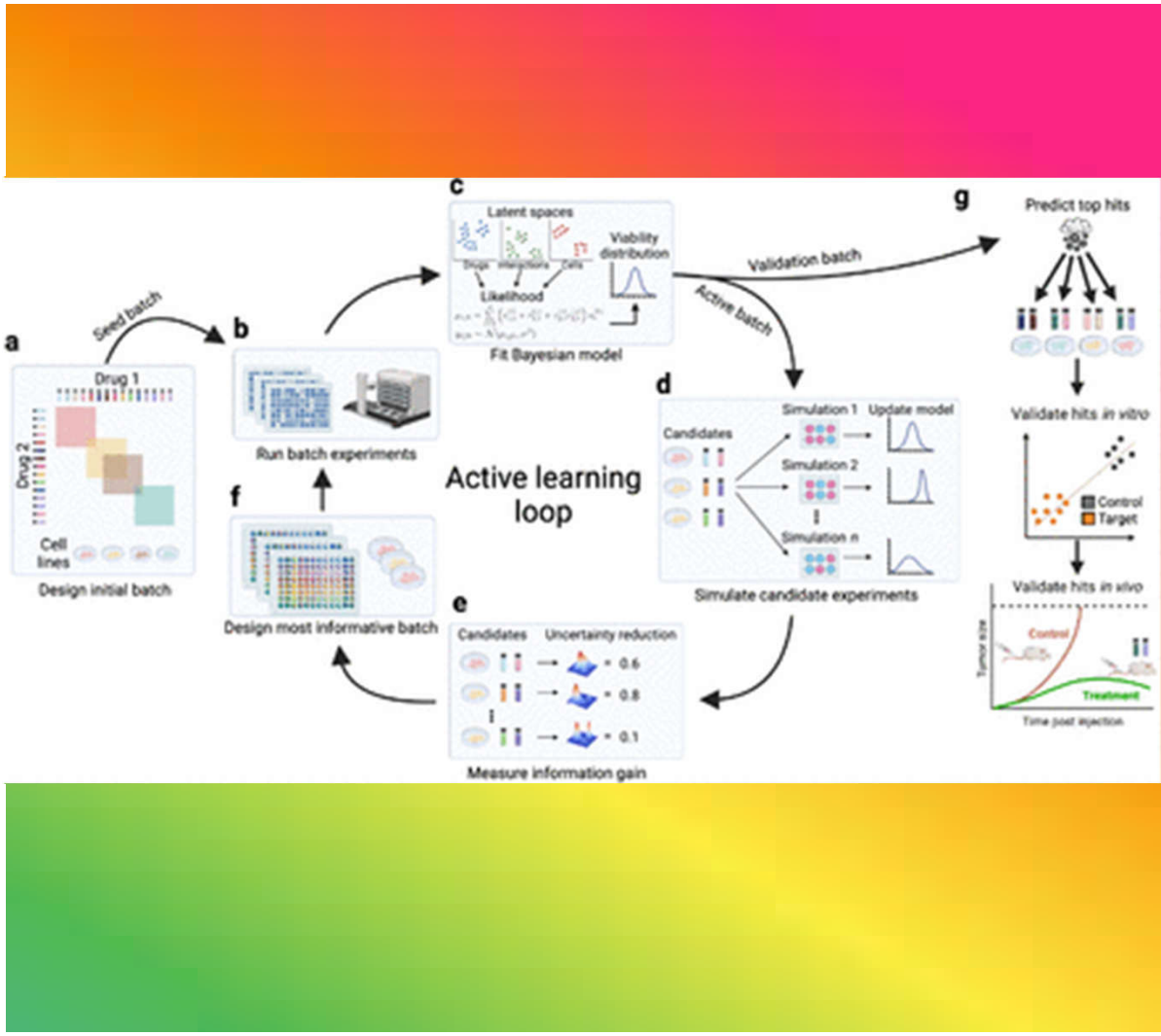


Applications for Bayesian Learning

Automatically Harvesting Broccoli

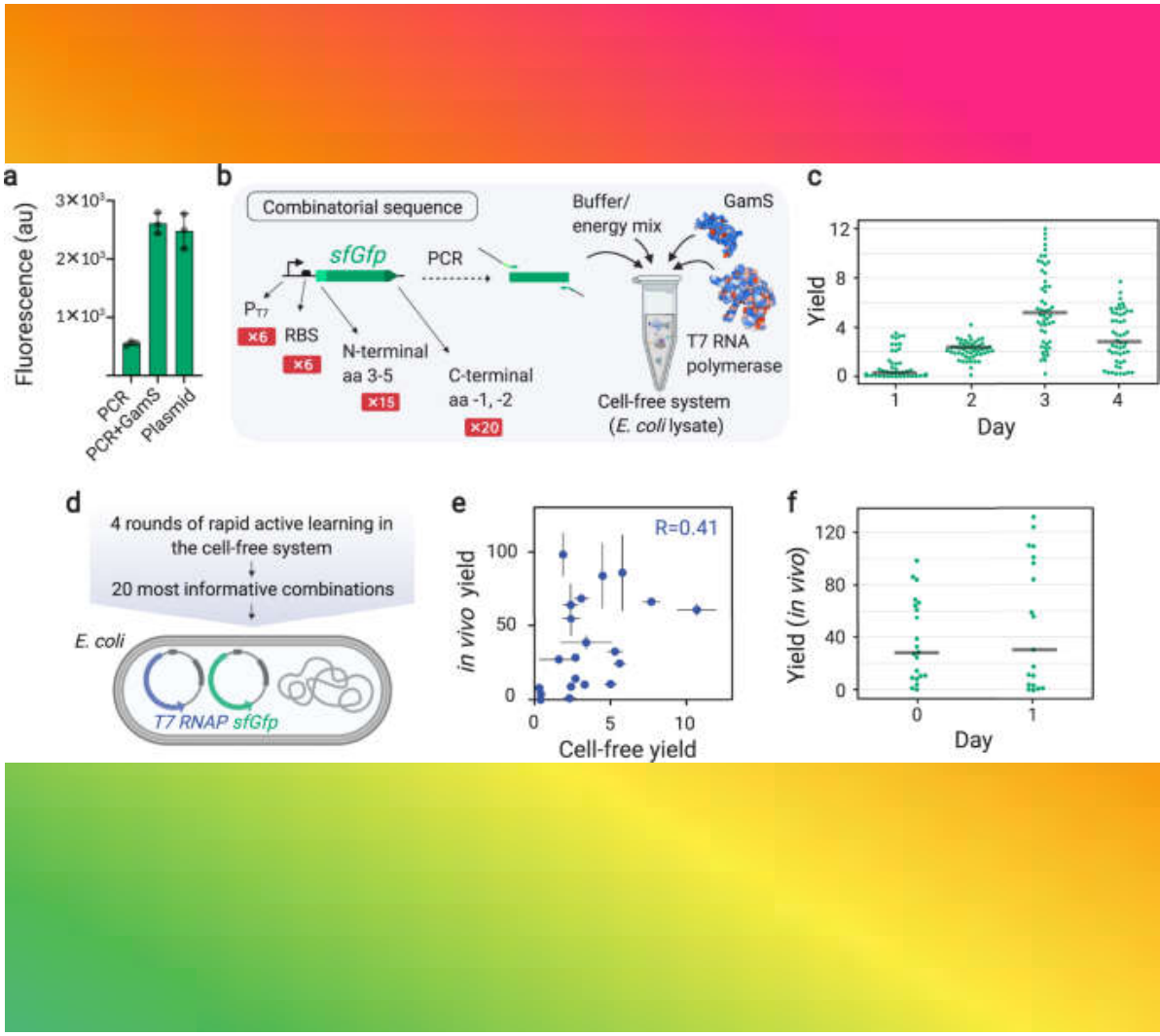
Applications for Bayesian Learning

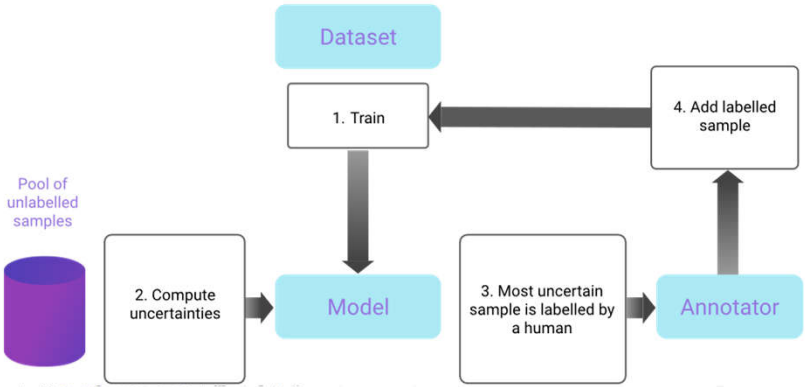
Researching Drug Combinations



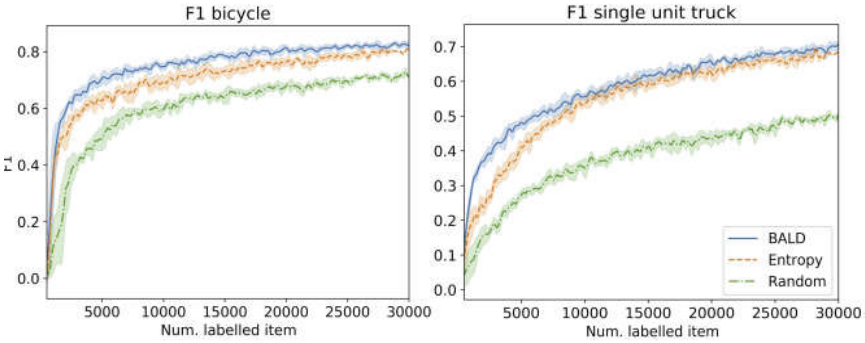
Applications for Bayesian Learning

Researching Genetic Sequencing





Our Experiment



We aimed to replicate the results of a paper generalizing Bayesian Batch algorithms as a package in which they found,

Performance of different active learning procedures on Mio-TCD. While any active learning method is strong against random, BALD is especially strong at the beginning of the labelling process. Performance averaged over 5 runs. (Atighehchian 2020)

Our experiment is still running on the Rose-Hulman GPU sever right now.

Paper mentioned over training problem with query size one early results indicate over training.

Significance & Extension

- Better Heuristic evaluation
- Focus on scalability
- Reduce model specificity
- Reduces biases