Ryerson University — Data Science & Analytics Graduate Studies

2018-2019 M.Sc. in Data Science and Analytics

Developing a Confidence Measure Based Evaluation Metric for Breast Cancer Screening Using Bayesian Neural Networks

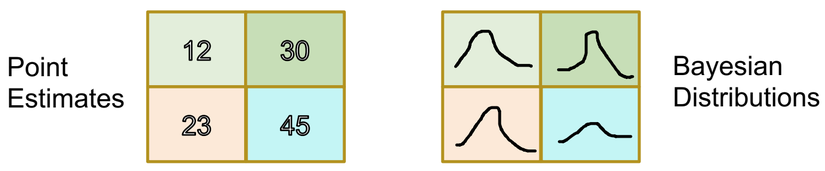
Anika Tabassum — **Supervisor: Dr. Naimul Khan**

Objective

- Coming up with a new evaluation criterion based on confidence measurement for breast cancer mammography image classification, so that in addition to classification accuracy, it provides a few numeric parameters that can be tuned to adjust the confidence level of the classification
- Leveraging Bayesian Deep Neural Networks and Transfer Learning for this purpose

Background

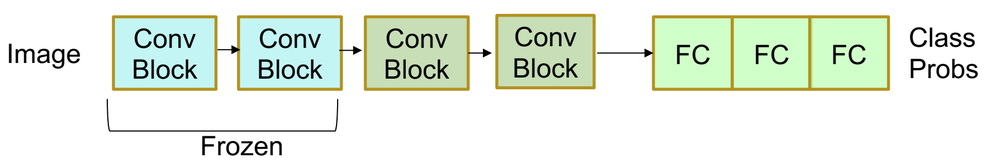
- Classification accuracy is a good metric; however, it does not provide any measure of confidence / uncertainty
- Crucial real world problems like mammography image screening require some confidence measure
- Bayesian approaches provide measure of uncertainty via its parameters in the form of probability distribution, rather than single point estimates



- Deep learning has been a wide success for image classification
- Bayesian Deep Neural Networks can be a good thing to explore for confidence / uncertainty measurement for images.

Methodology

- Train deterministic neural network
- Leverage Transfer Learning
- Pretrained ResNet-18 with some tweaks, replace last FC layer with 3 new FC, freeze first couple of blocks

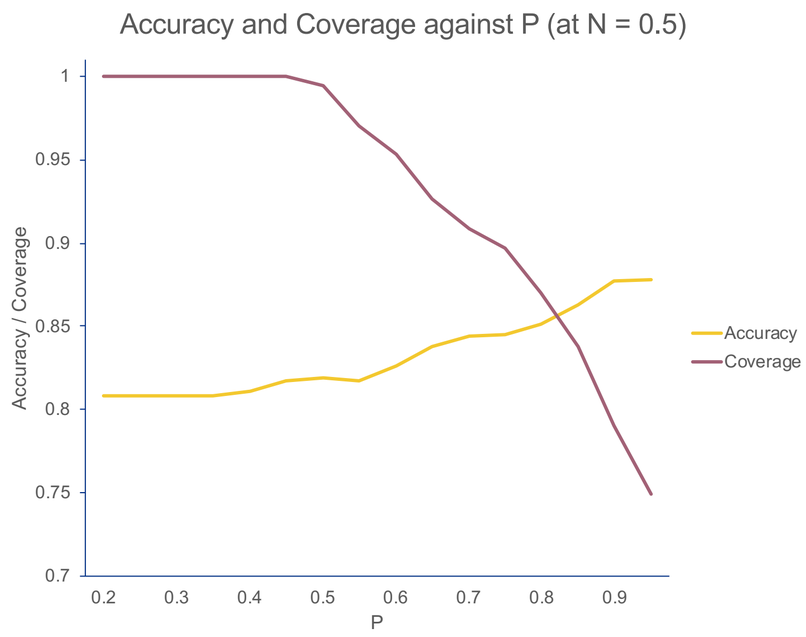


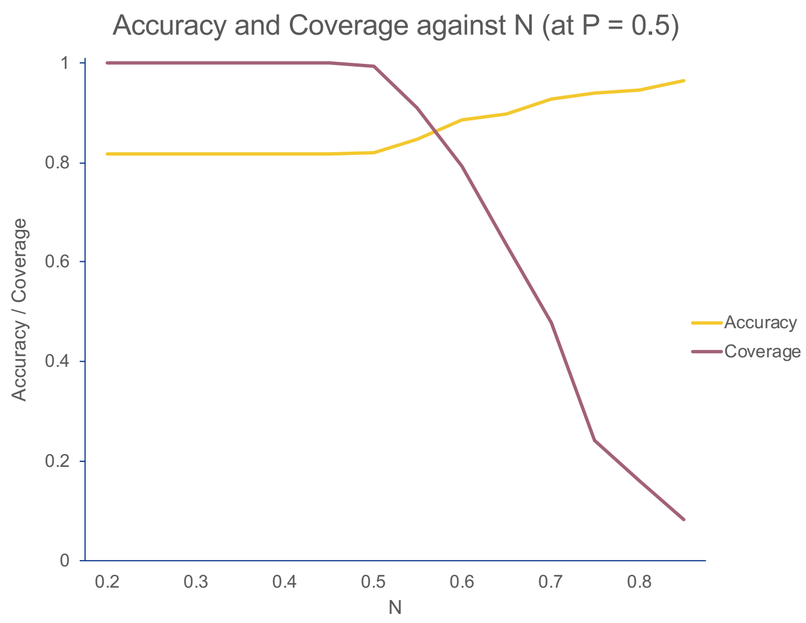
- Bayesian training for entire network too expensive
- Solution – extract the Conv blocks, use as feature generator only, not Bayesian layers
- Separate network with the FC layers – this is the Bayesian neural network

- Gaussian priors (0 mean, 1 std) and SVI (Stochastic Variational Inference) for posterior inference
- Sample 1000 deterministic networks from the posterior
- Two parameters N and P for confidence measure
- N = fraction of 1000 networks having a minimum prob for a class (benign or malignant) for an image
- P = that minimum probability
- Example: N = 0.7, P = 0.5 means at least 700 networks have to have minimum 0.5 prob for one of the (same) labels, otherwise skip the image (don't classify)

Results

- Increased value of P and N result in some images being skipped (lower coverage). However, accuracy on the non-skipped images increases
- N and P tunable for reaching a desired level in the accuracy-coverage trade-off





- A tuple (Accuracy, Coverage, N, P) can be our new evaluation criterion with (N, P) being the confidence measure.

Conclusions

- We have shown an approach that makes use of deep neural nets, transfer learning and Bayesian learning to come up with a confidence measure
- We have introduced some tunable parameters that allow adjusting the level of confidence / uncertainty
- Although done for breast cancer images and 2 classes, idea extendable to any domain and any number of classes
- Ample chance of future work in terms of network architecture selection, training strategy and confidence measurement