Uncertainty estimation in AI-assisted pathology with conformal prediction

Marco Capuccini, PhD student
Department of information technology
Department of pharmaceutical biosciences
Uppsala University (Sweden)
AI-assisted prostate cancer pathology with deep learning

~80K prostate biopsies from ~7.5K men

~80k slides

~5M training tiles

~1TB on disk

Train CNN with Inception-v3

Validation over ~500K tiles
Estimating uncertainty with conformal prediction

For a user-defined $\varepsilon$

$\text{CP}(\varepsilon, x_1) = \{\text{Benignant}\}$

$\text{CP}(\varepsilon, x_2) = \{\text{Benignant, Cancer}\}$

$\ldots$

$\text{CP}(\varepsilon, x_n) = \emptyset$

$x_k, k=1\ldots n$ is an unseen tile

By construction the true label of $x_k$ is in the prediction set with probability at least $1-\varepsilon$

(Vovk et al. provide proof under exchangeability assumption)
Conformal Predictor Efficiency

![Bar Chart](image.png)

Significance Level ($\varepsilon$)
Infrastructure

CNN training

- P2.2xlarge.8 x 1
- Ultra I/O block storage
- 60 minutes 1 epoch training (70% GPU usage on average)
- 20 minutes to run validation (30% GPU usage on average)
- CoreOS Container Linux

Conformal Prediction over 30 significance levels

- 1 x Master, 9 x Workers (32 vCPU, 64GB RAM each)
- CoreOS Container Linux
Questions?

marco.capuccini@it.uu.se

notebook here