Targeted sampling based on epistemic uncertainty for improving predictive performance of DL tasks in the case of costly data acquisition

Epistemic uncertainty is reducible uncertainty, meaning that it can (in theory)# be entirely eliminated by adding more training data to the model in the input spaces where epistemic uncertainty is high.

While this is a well-known fact, there exist few methods on how to identify suitable input spaces in a user-friendly manner to collect more training data.

In this thesis methodology to identify suitable areas of the input space should be developed to give easily understandable directions on how to collect more training data. This approach could be benchmarked to non-probabilistic approaches such as active learning. This is especially useful in applications where data acquisition is extremely costly (due to expensive equipment (SAR imagery), hazardous collection (PET-CD, MRI))

Steps:

(1) Quantify epistemic and aleatoric uncertainty in a relevant use case (preferably in a medical imaging application)
(2) Develop a theoretically-founded approach to identify target regions of the input space on where to collect more data (for instance, particular patients, timesteps, or geospatial regions)
(3) Measure improvement in predictive performance, model robustness, and reduction of epistemic uncertainty by adding more data based on the set of rules developed in step 2

Example use cases: medical imagery, radar imagery (SAR)

Literature:

Hüllermeier, Waegeman: Aleatoric and epistemic uncertainty in deep learning: an introduction to concepts and methods (2021)
https://link.springer.com/article/10.1007/s10994-021-05946-3

Nguyen et al., Epistemic Uncertainty Sampling (2019), https://arxiv.org/abs/1909.00218

Nath et. al., Diminishing Uncertainty within the Training Pool: Active Learning for Biomedical Image Segmentation (2021), https://arxiv.org/pdf/2101.02323.pdf