Technische Hochschule Ingolstadt

How to make good predictions

Jonas Ammeling

Common Questions

- How do we split the data?
- What is a good validation strategy?
- How do we know our model generalizes well to unseen data?

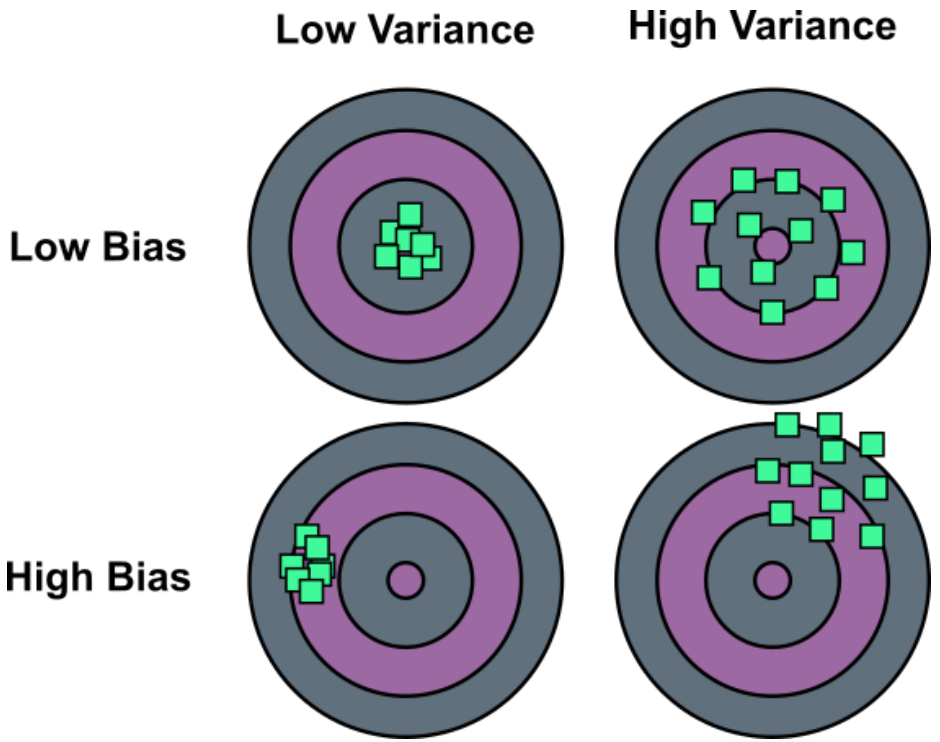
Pessimistic bias

If a model has not reached its capacity, the performance estimate on the test set would be pessimistically biased

Bias-Variance Tradeoff

- „The price to pay for achieving low bias is high variance“ (Geman 1992)

- Variance Decomposition (Neal et al. 1992:
  - Variance due to optimization
  - Variance due to sampling



The Holdout Method

- Simplest and most common evaluation technique

- Typically performed as 3-way holdout method
  - Training set for optimization
  - Validation set for hyperparameter tuning and model selection
  - Test set for evaluating predictive performance

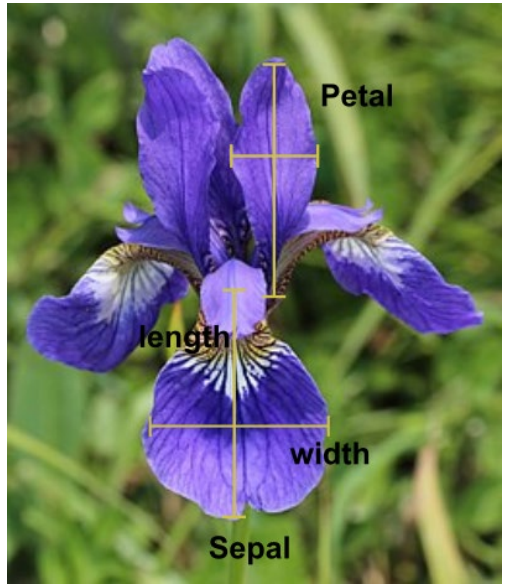
- Problems
  - Single performance estimate is subject to high bias and variance depending on the choice of data splits

Monte Carlo Cross-Validation

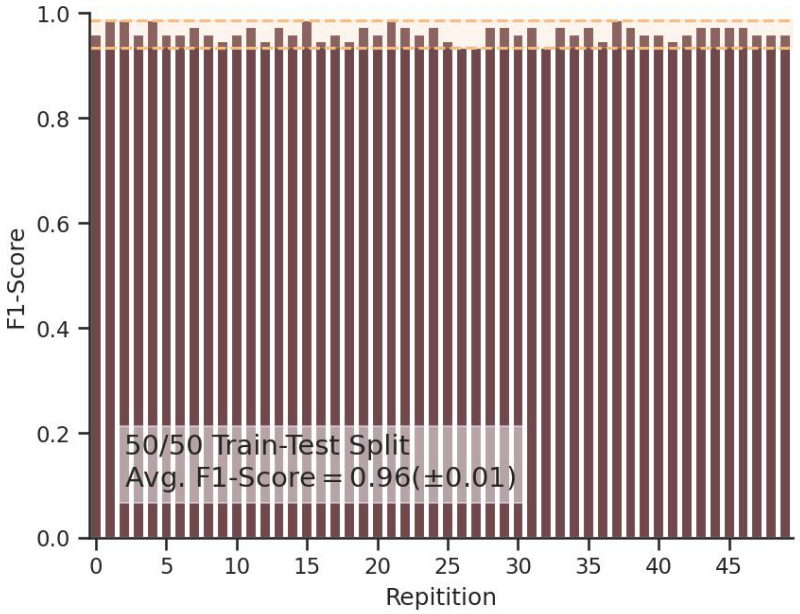
- Repeated holdout evaluation with random data splits

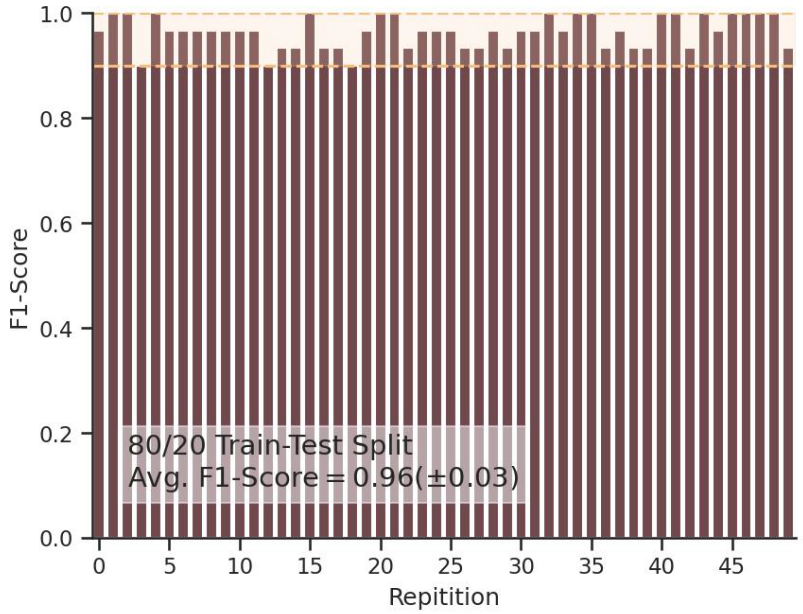
- More robust performance estimate
  - Averaging over multiple runs reduces variance due to sampling

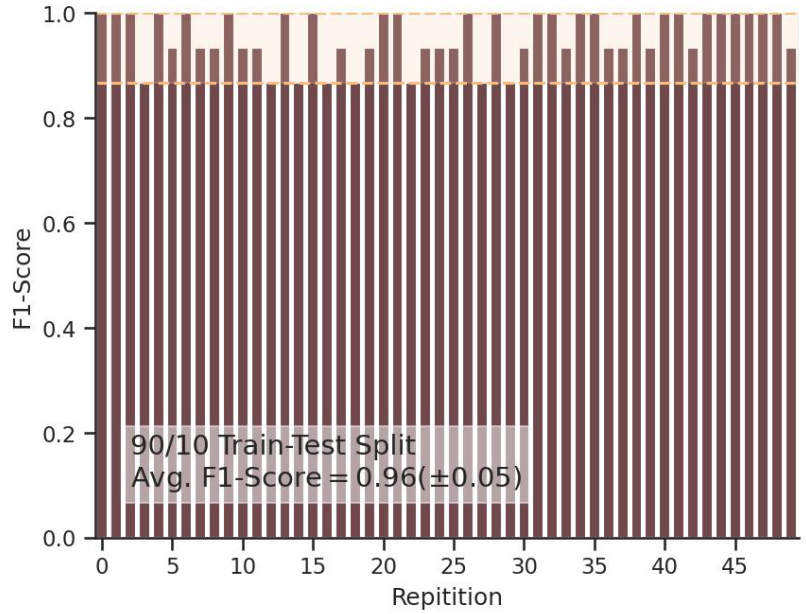
Monte Carlo Cross-Validation



- Model: 3-NN Classifier
- Dataset: IRIS (150 Instances)

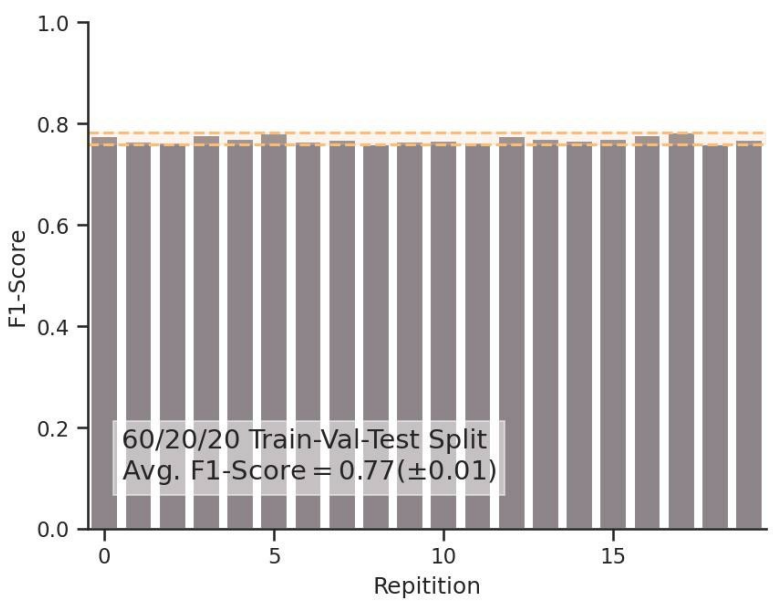


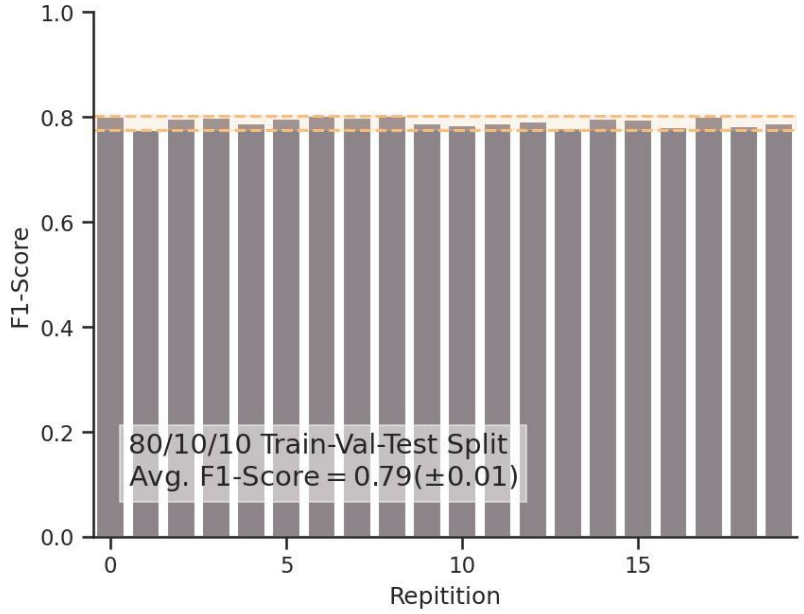


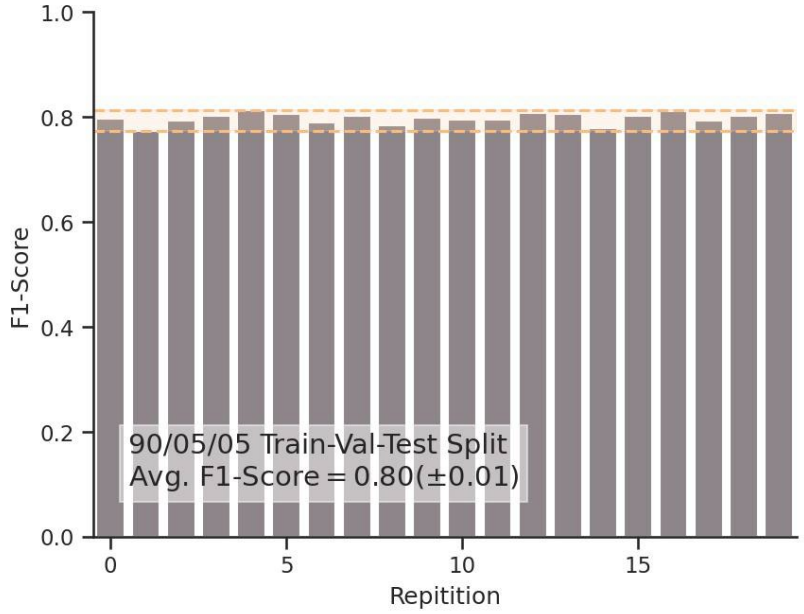


Monte Carlo Cross-Validation

- Model: Resnet18 (11.7 M)
- Data: CIFAR10 (60K Images)

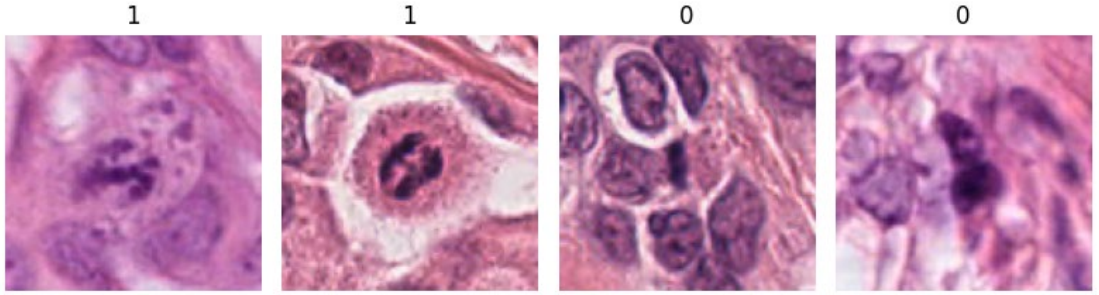


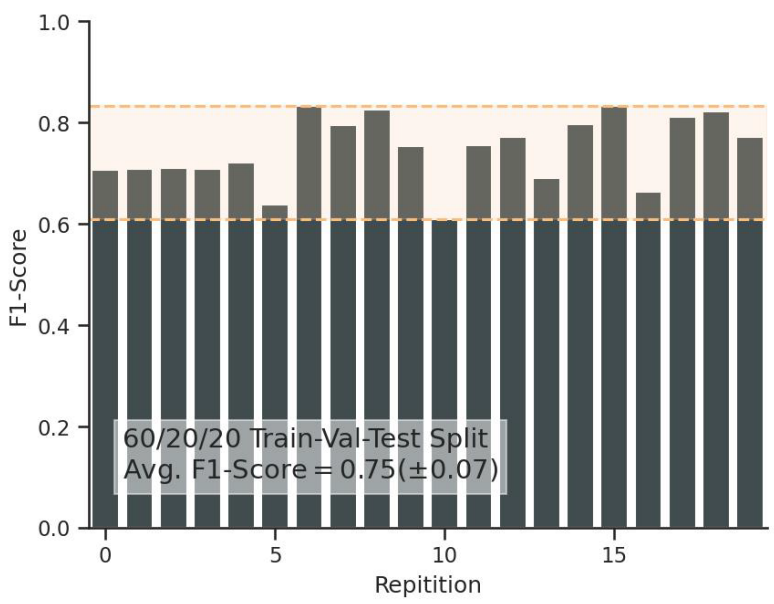


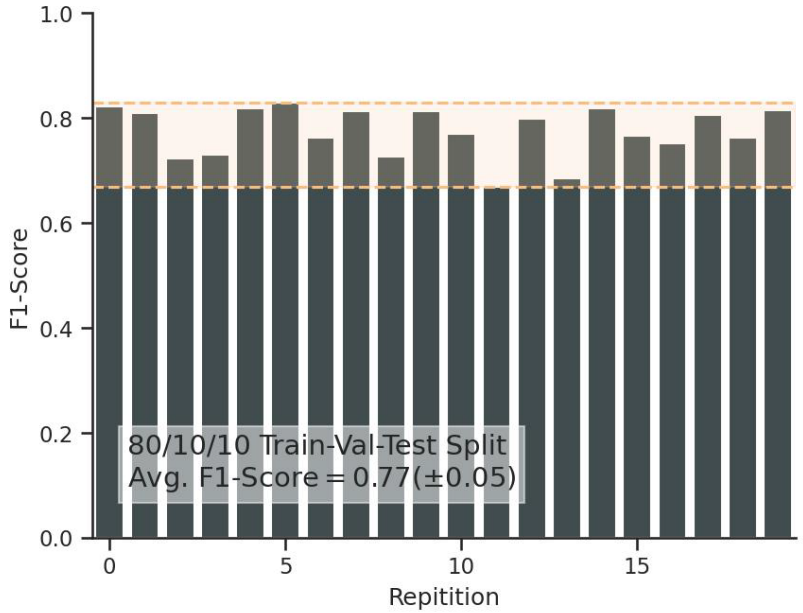


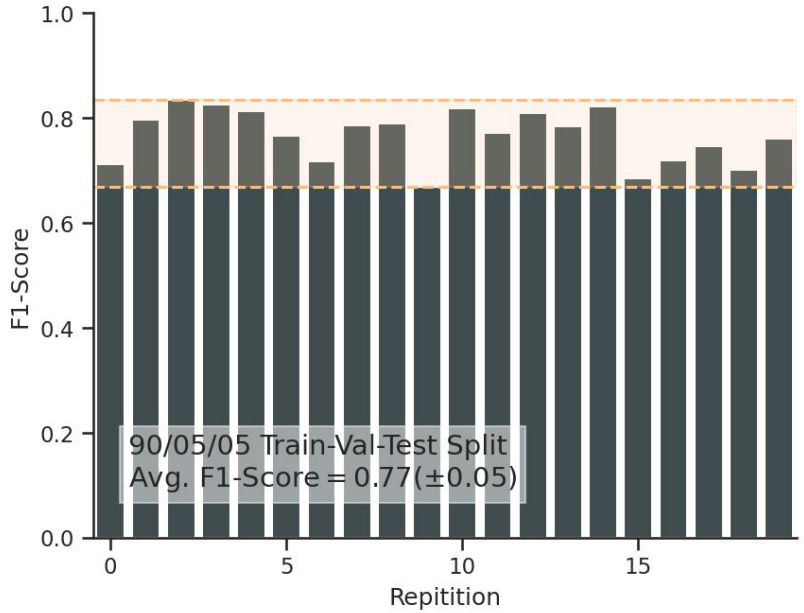
Monte Carlo Cross-Validation

- Model: Resnet18 (11.7 M)
- Data: MIDOG (150 Images, 1721 MFs)









Monte Carlo Cross-Validation

- Repeated holdout evaluation with random data splits

- More robust performance estimate
  - Averaging over multiple runs reduces variance due to sampling

- Problems
  - Some samples may never be part of the test set
  - (Performance estimates become dependent due to repeated use across repititions)

K-Fold Cross-Validation

- Most preferred technique for model evaluation and model selection
- Reduces pessimistic bias by using all samples for training and testing compared to standard holdout evaluation
- Test folds are non-overlapping compared to MCCV
- Guarantees that each sample is used for testing compared to MCCV

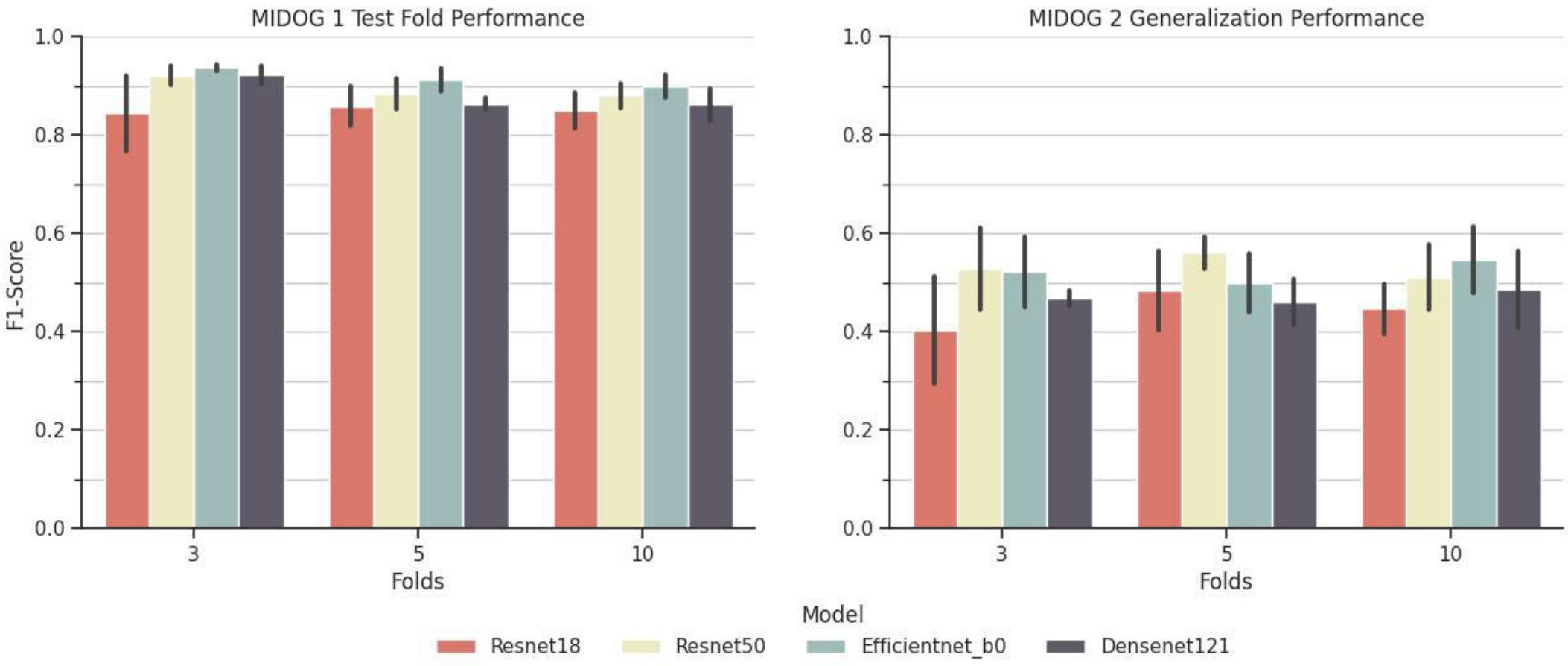
K-Fold Cross-Validation

- K-Fold CV on MIDOG for K = [3, 5, 10]

- Models
  - Resnet18 (11.7 M)
  - Resnet50 (25.6 M)
  - Efficientnet_b0 (5.3 M)
  - Densenet121 (8 M)

- External validation on MIDOG 2 (without hBC)

K-Fold Cross-Validation



Thank you!

References

- Geman, Stuart & Bienenstock, Elie & Doursat, René. (1992). Neural Networks and the Bias/Variance Dilemma. Neural Computation. 4.
- Neal, Brady & Mittal, Sarthak & Baratin, Aristide & Tantia, Vinayak & Scicluna, Matthew & Lacoste-Julien, Simon & Mitliagkas, Ioannis. (2018). A Modern Take on the Bias-Variance Tradeoff in Neural Networks.