Take-home messages

- Having **one dataset**, we need to be splitted into **training**, **validation** and **testing sub-cohorts**.
- Considering a part of the dataset, based on the **hold-one-out method**, for final testing of the trained model **prevents bias** in the trained model.
- **K-fold cross-validation** is recommended for **training ML models**.
- **Normalization** is an important step before training an ML model.
- In the normalization of the sub-cohorts, **apply** the same **mean and std extracted from training** sub-cohort to **the validation and test samples**.
- We cannot say generally **which ML classifier is the best**, we need to do training and validation for a specified dataset to select the best one.
- **Evaluation of the trained models** depends on the **application**, e.g. in the tumor classification having a model with lower **false negative rate** is of importance.