Generalization and Model Validation: A key reason for splitting datasets is to ensure that models can generalize well to unseen data. The training set is used for learning, the validation set helps in tuning hyperparameters and preventing overfitting, and the test set provides an unbiased evaluation of the final model. Without these splits, a model is at risk of overfitting, learning only the noise in the training data, and failing to generalize to new, real-world scenarios​. :[Dataset split](https://ar5iv.labs.arxiv.org/html/2106.04525)