**Highlights**

- Model evaluation reliability is determined by the choice of cross-validation strategy.

- Misuse of model selection can lead to over-optimistic performance estimates.

- Overlooking experimental block effects can lead to biased model performance estimates.

- Regression metrics can be misleading when the data structure is complex.

- Classification metrics can be misleading when the class distribution is imbalanced.