Pseudocode for Model Assessment and Model Selection

Algorithm 1 Model Assessment (k-fold CV)

Require: Dataset D, set of configurations Θ, number of repetitions R

for $r \leftarrow 1$ to R **do**
  Shuffle D
  Split D into k folds $F_1, \ldots, F_k$
  for $i \leftarrow 1$ to k **do**
    $train_k, test_k \leftarrow (\bigcup_{j \neq i} F_j), F_i$
    $best_{model} \leftarrow \text{Select}(train_k, \Theta)$
    $perf_k \leftarrow \text{Eval}(best_{model}, test_k)$
  end for
  $perf_r \leftarrow \sum_{i=1}^{k} perf_i / k$
end for
$perf \leftarrow \sum_{r=1}^{R} perf_r / R$
return $perf$

Algorithm 2 Model Selection

Require: $train_k$, Θ

$valid \leftarrow^{R} 10\% * train_k$
$train \leftarrow train_k \setminus valid$
$best_{model} \leftarrow$ None
$best_{perf} \leftarrow -\infty$
for all $\theta \in \Theta$ **do**
  $model \leftarrow \text{Train}(train, \theta)$
  $perf \leftarrow \text{Eval}(model, valid)$
  if $perf > best_{perf}$ **then**
    $best_{perf} \leftarrow perf$
    $best_{model} \leftarrow model$
  end if
end for
return $best_{model}$

In Algorithm 1, "Select" refers to Algorithm 2. "Train" and "Eval" represent training and testing phases respectively. After each model selection, the best model is used to evaluate the external test fold. Performances are averaged across the k folds and R repetitions of the cross validation.