

Applicability Domain Evaluation for the Validation Set

For each compound in the validation set, the Tanimoto similarity to all training compounds was calculated using the same fingerprint representation. The maximum similarity value was retained for each validation compound, corresponding to its nearest neighbor in the training set. All validation compounds were classified as being outside the domain. The query-based nearest-neighbor approach ensures that each validation compound is evaluated individually against the chemical space defined by the training set and avoids any overlap between the training set and the validation data.

