Mode 1: Rayyan screening

Updated set (N=1,126)

Load in Rayyan

Manual screening

Relevant (+)

Irrelevant (-)

Mode 2: AIM Review screening

Updated set (N=1,126)

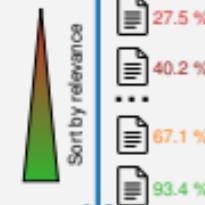Load in AIM Review

Manual screening

Relevant (+)

Irrelevant (-)
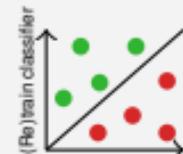
Mode 3: AIM Review screening with active learning

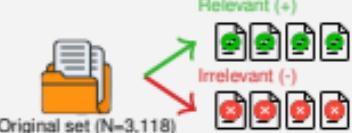Updated set (N=1,126)

Predict relevance scores

Sort by relevance

27.5 %
40.2 %
67.1 %
93.4 %

Cycle until finish

Manual screening

Relevant (+)

Irrelevant (-)

(Re)train classifier
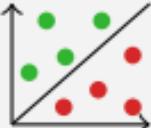
Original set (N=3,118)
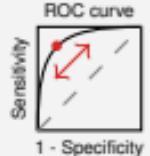
Mode 4: AIM Review automatic screening

Original set (N=3,118)

Relevant (+)

Irrelevant (-)

Nested cross-validated classifier development

Probability threshold selection

ROC curve

Sensitivity

1 - Specificity

Updated set (N=1,126)

Predict unlabeled