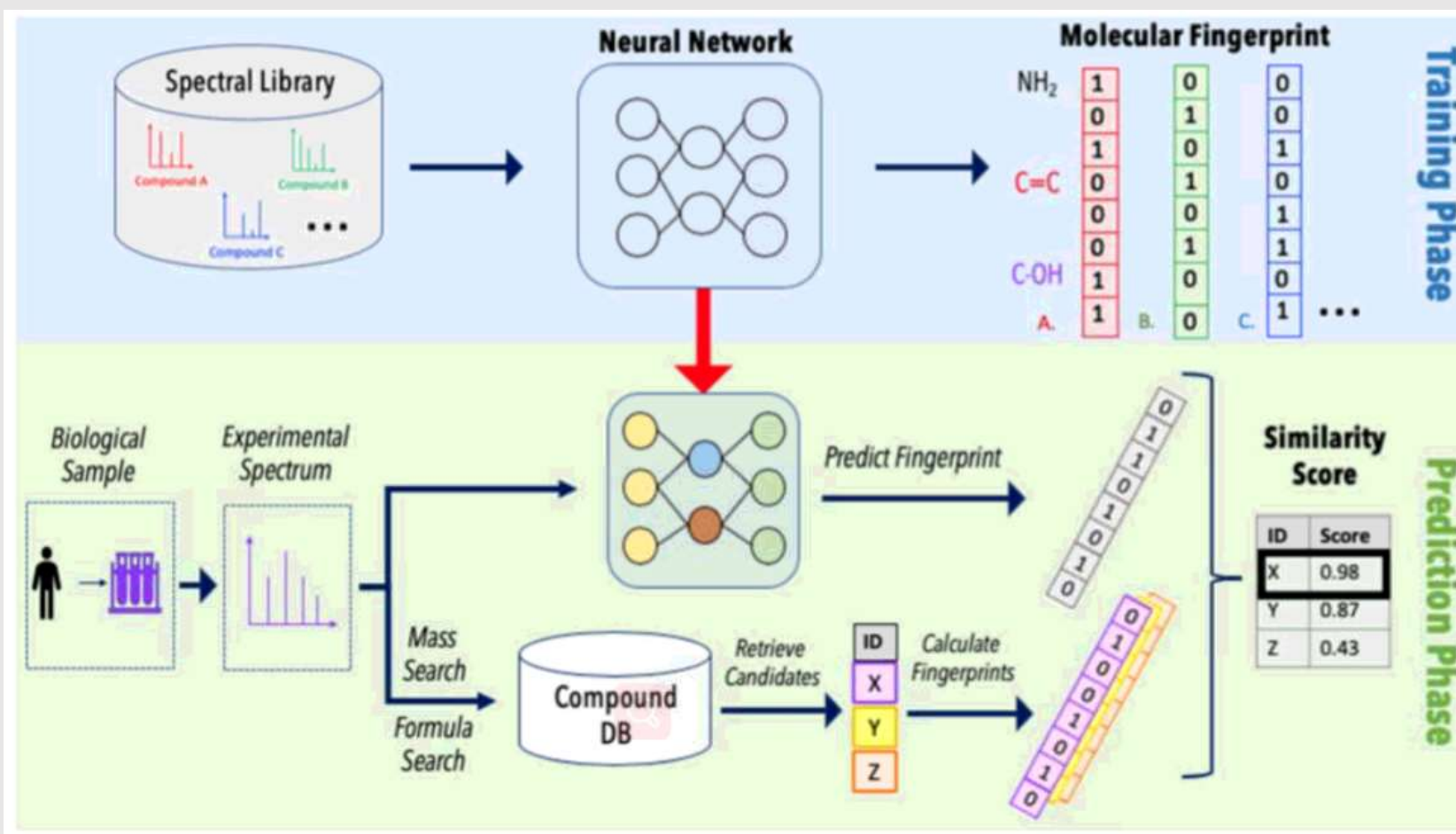



Fig. 2 Steps involved in processing spectra prior to training an ANN



Merging



Fig. 2 Steps involved in processing spectra prior to training an ANN

Input Data	Type	Size
NIST20_LR_IT_H_n.p	DataFrame	11147x19
NIST20_LR_IT_H_p.p	DataFrame	24377x19
NIST20_LR_IT_Na.p	DataFrame	5578x19
NIST20_LR_QQQ_H_n.p	DataFrame	929x19
NIST20_LR_QQQ_H_p.p	DataFrame	18750x19
NIST20_LR_QQQ_Na.p	DataFrame	2080x19

fingerprint.py

NIST_LR.p

vector_transformation.py

NIST_LR_vec.p

fp_conversion.py

NIST_LR_fp.p

NIST_LR_fp.p Dict 26006 169,808KB

fingerprint.py

NIST_HR.p

vector_transformation.py

NIST_HR_vec.p

fp_conversion.py

NIST_HR_fp.p

NIST_HR_fp.p Dict 25550 180,083KB

Input Data	Type	Size
NIST20_HR_HCD_H_n.p	DataFrame	112023x19
NIST20_HR_HCD_H_p.p	DataFrame	266321x19
NIST20_HR_HCD_Na.p	DataFrame	33099x19
NIST20_HR_ITFT_H_n.p	DataFrame	10377x19
NIST20_HR_ITFT_H_p.p	DataFrame	20738x19
NIST20_HR_ITFT_Na.p	DataFrame	5163x19
NIST20_HR_QTOF_H_n.p	DataFrame	2419x19
NIST20_HR_QTOF_H_p.p	DataFrame	24601x19
NIST20_HR_QTOF_Na.p	DataFrame	1278x19

Prediction-model training

NIST20

NIST Mass spectra
(LR945, HR986)

relation

NIST Fingerprint(528)

Train set
NIST Mass spectra
(LR945, HR986)

ANN model 1

Train set
NIST Fingerprint(528)

Test set
NIST Mass spectra
(LR945, HR986)

ANN model 1

Test set
NIST Fingerprint(528)

Test set
predict Fingerprint(528)

Input data: NIST_LR_fp.p / NIST_HR_fp.p

Code: model_evaluation.py / model_evaluationHR.py

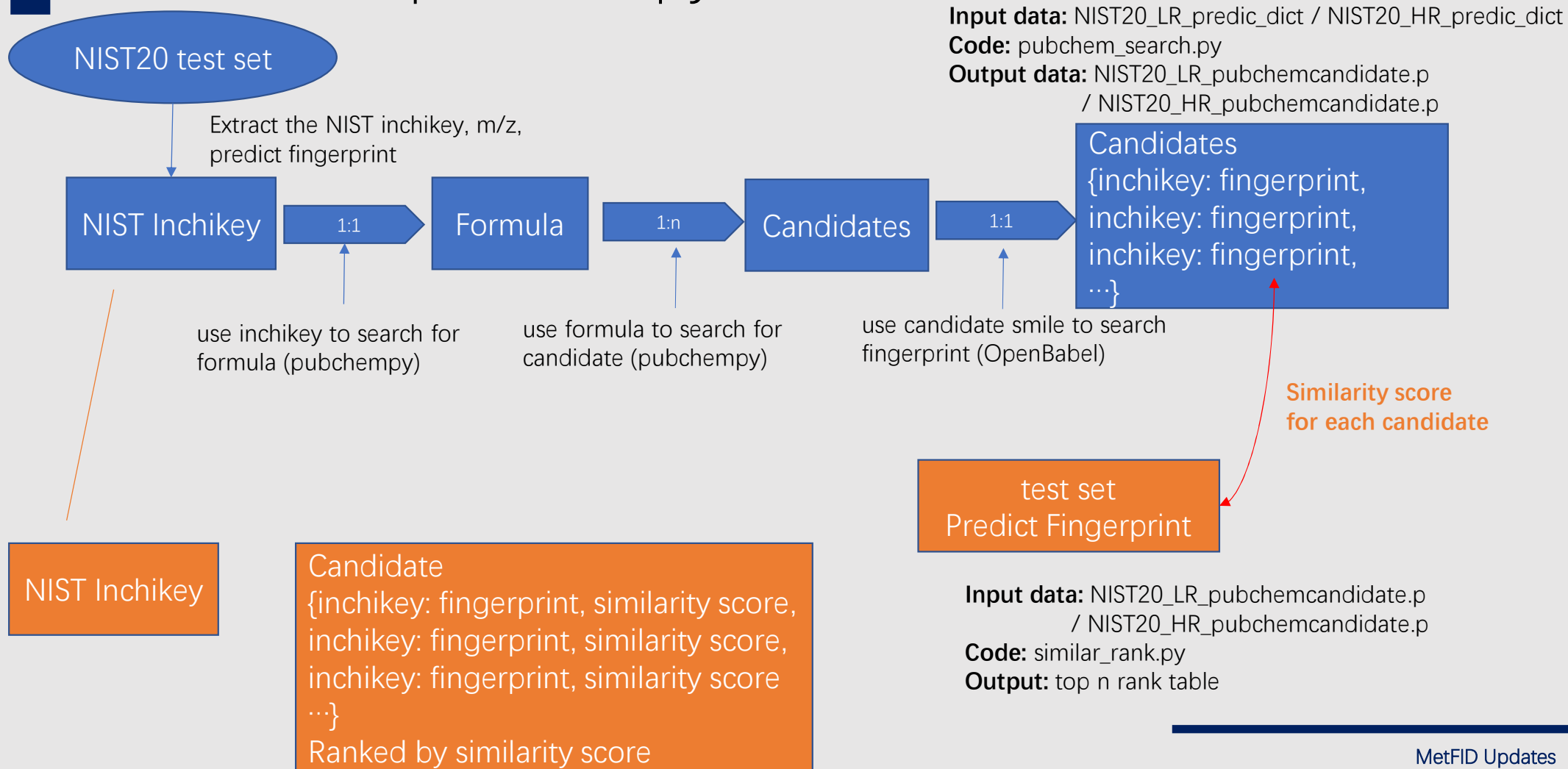
Output data: NIST20_LR_predic_dict / NIST20_HR_predic_dict



Fig. 2 Steps involved in processing spectra prior to training an ANN

Model	input	train vs test	binary accuracy
ANN	NIST20_LR	21473 vs 4533	92.43%
ANN	NIST20_HR	21017 vs 4533	93.93%

Evaluation – pubchempy search



Evaluation – omic_craft search

