Supporting Information: Machine Learning anomaly detection of automated HPLC experiments in the Cloud Laboratory — Supporting Information

Filipp Gusev,[†,‡] Benjamin C. Kline,[¶] Ryan Quinn,[§] Anqin Xu,[¶] Ben Smith,[¶] Brian Frezza,[¶,§] and Olexandr Isayev[*,†,‡]

†*Department of Chemistry, Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA, 15213*

‡*Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 15213*

¶*Emerald Cloud Lab, 15500 Wells Port Dr, Austin, TX 78728, USA*

§*Emerald Therapeutics, 15500 Wells Port Dr, Austin, TX 78728, USA*

E-mail: olexandr@olexandrisayev.com

Contents

List of Figures

List of Tables

SI-1 SI Methods

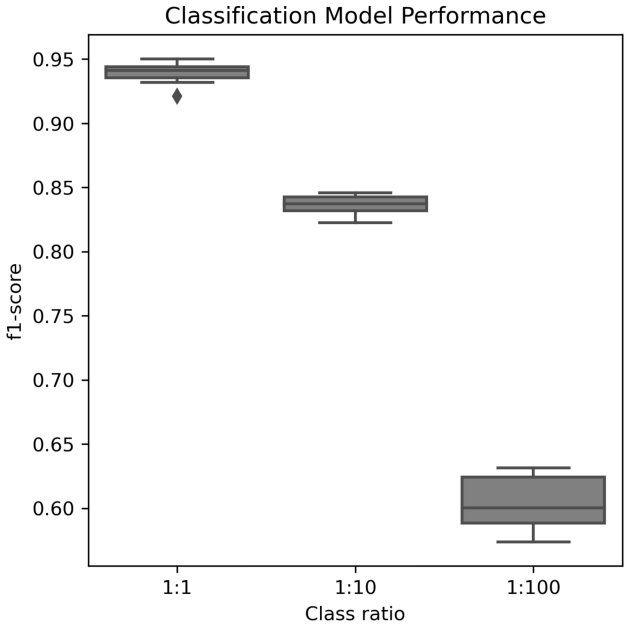

Figure S1: Classification ML-model performance assessment for selecting target class ratio. F1-score measured across 5 stratified cross-validation splits.

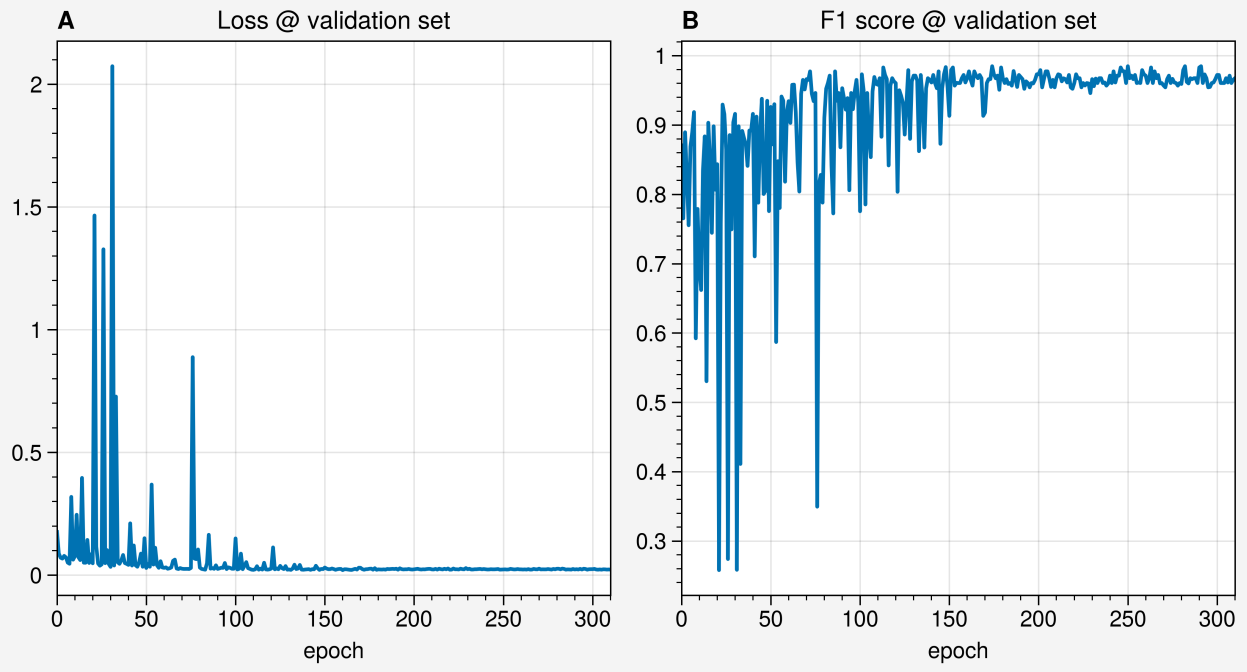

Figure S2: CNN model training performance at final annotation round.

Table S1: CNN model hyperparameters for Round 2 Model and Deployment Model.

Parameter	**Description**	**Search Range / Values**	**Round 2 Model**	**Deployment Model**
MAX_SEQ_LEN	Maximum sequence length	Fixed at 1000	1000	1000
SKIP_FIRST_N	Number of initial elements to skip in sequence	Integer: 0 to 60, step 5	0	40
BATCH_SIZE	Batch size for training	Fixed at 100	100	100
n_feature_layers	Number of feature extraction layers	Integer: 1 to 7	3	3
n_feature_layers_och	Output channels for each feature layer	List of integers: 1 to 12, step 1	[2,4,8]	[11,11,9]
n_feature_layers_ich	Input channels for each feature layer	Derived from 'n_feature_layers_och'	[1,2,4]	[1,11,11]
n_feature_layers_ksize	Kernel size for each feature layer	List of integers: 1 to 9, step 1	[5,3,3]	[8,4,6]
n_feature_layers_stride	Stride for each feature layer	List of integers: 1 to 9, step 1	[3,3,3]	[4,5,4]
n_agg_layers	Number of aggregation layers	Integer: 1 to 3	2	1
n_agg_layers_och	Output channels for each aggregation layer	Log-scaled integers: 8 to 128	[288,64]	[25]
lr_init	Initial learning rate	Log-scaled float: 1e-7 to 1e-2	1e-5	5.8e-4

Table S2: Number of HPLC experiments in initial dataset per Chromatography Type, Instrument Model Name, and Instrument Manufacturer

Chromatography Type	Instrument Model Name	Instrument Manufacturer	Count
ReversePhase	UltiMate 3000	Dionex	15622
SizeExclusion	UltiMate 3000	Dionex	5867
IonExchange	UltiMate 3000	Dionex	2196
ReversePhase	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	598
ReversePhase	UltiMate 3000 with FLR Detector	Dionex	312
SizeExclusion	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	231
SizeExclusion	UltiMate 3000 with FLR Detector	Dionex	163
IonExchange	UltiMate 3000 with FLR Detector	Dionex	144
ReversePhase	UltiMate 3000 with PCM Detector	Dionex	88
ReversePhase	Waters Acquity UPLC H-Class FLR	Waters	85
ReversePhase	Waters Acquity UPLC H-Class PDA	Waters	41
IonExchange	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	35
ReversePhase	Waters Acquity UPLC H-Class ELS	Waters	31
IonExchange	Waters Acquity UPLC H-Class PDA	Waters	7
IonExchange	Waters Acquity UPLC H-Class FLR	Waters	3

Table S3: Dataset for final model: Count of samples by instrument characteristics and annotation type

Chromatography Type	Instrument Model Name	Instrument Manufacturer	Expert Annotation AirBubble	Expert Annotation Normal	SNA Annotation Normal
IonExchange	UltiMate 3000	Dionex	78	24	558
IonExchange	UltiMate 3000 with FLR Detector	Dionex	45	0	2
IonExchange	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	0	14	1
IonExchange	Waters Acquity UPLC H-Class PDA	Waters	0	2	0
ReversePhase	UltiMate 3000	Dionex	343	112	4290
ReversePhase	UltiMate 3000 with FLR Detector	Dionex	2	5	70
ReversePhase	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	45	85	57
ReversePhase	UltiMate 3000 with PCM Detector	Dionex	4	8	5
ReversePhase	Waters Acquity UPLC H-Class ELS	Waters	0	0	13
ReversePhase	Waters Acquity UPLC H-Class FLR	Waters	0	0	19
ReversePhase	Waters Acquity UPLC H-Class PDA	Waters	0	5	7
SizeExclusion	UltiMate 3000	Dionex	177	1	1604
SizeExclusion	UltiMate 3000 with FLR Detector	Dionex	5	5	47
SizeExclusion	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	1	0	66
Total			700	261	6739

Table S4: Performance of the model on the prospective validation set, grouped by Chromatography Type and Instrument Model

Chromatography Type	Instrument Model Name	Instrument Manufacturer	True Positives	True Negatives	False Positives	False Negatives	Total	Accuracy
IonExchange	UltiMate 3000	Dionex	65	357	2	4	428	0.99
IonExchange	UltiMate 3000 with MALS-DLS-RI Detector	Dionex	14	8	0	13	35	0.63
IonExchange	Waters Acquity UPLC H-Class PDA	Waters	0	25	0	1	26	0.96
ReversePhase	UltiMate 3000	Dionex	76	199	16	1	292	0.94
ReversePhase	UltiMate 3000 with FLR Detector	Dionex	0	9	0	0	9	1.00
ReversePhase	Waters Acquity UPLC H-Class FLR	Waters	0	11	0	0	11	1.00
SizeExclusion	UltiMate 3000	Dionex	82	82	2	0	166	0.99

Table S5: Chromatographic Operating Space: Parameter ranges used in the study

Chromatography Parameter	**Parameter Range**
Separation Mode	{ReversePhase; IonExchange; SizeExclusion}
Chromatography Scale	{Preparative; Analytical}
Column	{Aeris XB-C18 Peptide Column 10mm x 150mm; XBridge Peptide BEH C18 300 Angstrom Column 150mm; XBridge Peptide BEH C18 150mm Prep Column; XBridge Oligonucleotide BEH C18 4.6x50mm; DNAPac PA200 4x250mm Column; XBridge Peptide BEH C18 300 Angstrom Column 150mm - NO PYRIDINE; DNAPac PA200 9x250mm Column; XBridge Protein BEH SEC 125 Angstrom Column 150mm; Aeris XB-C18 15cm Peptide Column; Aeris WIDEPORE XB-C8 150x4.6mm 3.6um particle 20nm pore Column}
Column Temperatures	{5°C; 15°C; 22°C; 25°C; 35°C; 37°C; 40°C; 45°C; 50°C; 55°C; 60°C}
Buffer A	{0.1M Triethylammonium Acetate; Ion Exchange Buffer A Native; Reverse phase buffer A 0.1% TFA; Milli-Q water; Filtered PBS, Non-Sterile; 15mM TEA 400mM HFIP; 0.1 M Ammonium Acetate}
Buffer B	{0.1M Triethylammonium Acetate (50% Acetonitrile); Ion Exchange Buffer B Native; Reverse phase buffer B 0.1% TFA; Milli-Q water; Filtered PBS, Non-Sterile; 15mM TEA 400mM HFIP diluted 50% MeOH; 0.1 M Ammonium Acetate (50% MeCN)}
Total Run Time	{17.5 min; 20.5 min; 21 min; 22 min; 25 min; 25.5 min; 28 min; 30 min; 30.5 min; 33 min; 34 min; 35.5 min; 38 min; 40 min; 43 min; 43 min; 45 min; 47.5 min; 50 min; 59 min; 60 min; 65 min; 70 min}
Flow Rate	{1 mL/min; 1.4 mL/min; 2.5 mL/min}