Demo of MLP and UMAP combined on CYTOF-Panorama samples

Version 1.1, March 2022

This demo teaches how to build a training set (classifier) for AutoGate's **Multi-Layer Perceptron (MLP)** with **Uniform Manifold Approximation Method (UMAP)**, based on the classifications done by the biologists (manual gates) in **Cytof Panoroma's published gating strategy** - <u>http://cgworkspace.cytogenie.org/Tutorials/omipB.png</u>

For reference, full publication is available here https://www.nature.com/articles/nmeth.3863?WT.feed_name=subjects_haematopoiesis

This document describes how to build a MLP model using the manual gates and then apply the model on other samples.

1 Set up the experiment

1.1 Click "Open AutoGate demo experiment" on the main window



AutoGate responds by showing you the below list of demo experiments available as of February 2022

	10 available demos				
	Select a demo and click the <mark>Ok</mark> bu (To see supporting documentation click buttons Publicat	itton ion, Gating or Tutorial))		
				1	8
	Basophil Activation Test 6 stains by Yael Gamez, Richard B. Moss et al	Publication		Tutorial	
	Macrophages and B cells 10 stains by Eliver Ghosn, Leonore Herzenberg et al	Publication	Gating	Tutorial	
	Pre-gated Macrophages and B cells 10 stains by Eliver Ghosn, Leonore Herzenberg et al	Publication	Gating	Tutorial	
	• CyTOF Panorama data set 49 stains by Nikolay Samusik, Garry P Nolan et al	Publication	Gating	Tutorial	
\boldsymbol{k}	OMIP-069: Forty-Color Full Spectrum Flow 40 stains by Lily M. Park et al@Cytek Biosciences Inc.	Publication	Gating	Tutorial	
	CyTOF Quantitative Comparison 50 stains by William E. O'Gorman et al@Genentech Inc.	Publication	Gating	- 1	
	OMIP-047: Hi-D Phenotypic Characterization of B Cells 16 stains by Thomas Liechti, et al.	Publication	Gating	- 1	
	OMIP-044: 28-Color Immunophenotyping Human Dendritic 28 stains by Florian Mair & Martin Prlic	Publication	Gating	-	
	Comparison of CyTOF assays across sites 22 stains by Michael D. Leipold, Holden T. Maecker et al	Publication	Gating	-	
	OMIP-044 with EPP: 28-Color Immunophenotyping Human Dendritic 28 stains by Florian Mair & Martin Prilic	Publication	Gating	-	
				× Cancel	Ok

1.2 Pick CytoF Panorama data set

AutoGate responds by opening the GatingTree window on the publication's fully stained samples. The bundled demo has already completed the setup plus the replication of the published manual gating (indicated by \mathbb{M}) hierarchy for one sample and being applied (indicated by \mathbb{A}) on rest of the samples.



Note: If you have already opened the panorama dataset and changed the GatingTree, you can start again using **Reset analysis to original demo** from the **File** menu



- 2 Build a training set(classifier) in supervised classification mode for both UMAP and MLP, based on the classifications (manual gates) done by the experts.
- 2.1 Click Hi-D from the menu and then MLP neural network under Configure/train



2.2 AutoGate responds by bringing up the following MLP configuration screen.

Train MLP on	"Manual gates"
	Train on gating sequence's
•	Leaf gates only Leaf or last branch gates
Specify reagents/parameters Pick using f or	(29) or Q X
 Python's TensorFlov MATLAB's fitcnet 	v Epochs limit: 500
Add to the training 0/9 other related gates	See in UMAP
(Cancel Train

Note:

- 1) **Python** is required to run TensorFlow on the computer, which can be downloaded here <u>https://www.python.org/downloads/</u>
- To use MATLAB's fitcnet code base, it is required that AutoGate is run on Matlab R2021a or later version. To download, click here - <u>https://au.mathworks.com/products/compiler/matlab-runtime.html</u> This tutorial is documented on Matlab R2021b

LP code base	
O Python's TensorFlow	Iteration limit: 1250
O MATLAB's fitcnet	% hold out 25 for validating
Add to the training 0/9 other related gates	See in UMAP
Can	cel Train

2.3 Specify the parameter for MLP to train on

To effectively build a training set based on the published manual results for a specific sub hierarchy we need to identify both the parameters that

- 1. Were used for the manual analysis
- 2. Those that look informative for the same cell types

2.4 Click on the tree icon to point MLP at parameters used for **Manual gates**

Train MLP on "N	lanual gates"
ſŢ	rain on gating sequence's
	Leaf gates only Leaf or last branch gates
Specify reagents/parameters (25 Pick using The pr m of MLP code base Pick reagents fr)) r Q X
Python's TensorFlow	Iteration limit: 1250
O MATLAB's fitcnet	% hold out for validating
Add to the training 0/9 other related gates	See in UMAP
Ca	ncel Train

2.4.1 Pick Manual gates of the first sample in the gating tree and click Pick

GatingTree for associated reagents picking	
Q Enter GatingTree search terms ▼	
All samples All samples Black of #1 Black of the second	
✓ ◆ 🕅 Non-Eosinophils<66,321/71,929 92%>	
♦ Plasma Cells<118/66,321 0.18%>	
✓ ◆ ☐ non-Plasma Cells<66,096/66,321 99.66%> ¹	
♦ pDCs<3,136/66,096 4.7%>	
<u>M</u>	
→	
BM2_cct_normalized_02_non-Neutrophils.fcs <87,772 events> A	
✓ ◆ ♣ Manual gates<87,772/87,772 100%> ▲	
✓ ♦ 🕅 non-Neutrophils<75,453/87,772 86%> 🗠	
Eosinophils<4,991/75,453 6.6%>	
> 🔷 🏧 Non-Eosinophils<69,998/75,453 93%> 🗠	
BM2_cct_normalized_03_non-Neutrophils.fcs <84,556 events>	
BM2_cct_normalized_04_non-Neutrophils.fcs <80,713 events>	
BM2_cct_normalized_05_non-Neutrophils.fcs <84,610 events>	
BM2_cct_normalized_06_non-Neutrophils.fcs <77,282 events> A	
BM2_cct_normalized_U/_non-Neutrophils.fcs <95,087 events> IA	
The second second of the local second second second second second	× Cancel

AutoGate returns to the configuration window. The parameters can also be specified using the Reagent Table.

Train MLP o	on "Manual gates"
	Train on gating sequence's
	Leaf gates only Leaf or last branch gates
Specify reagents/parameters	(29) or Q X
MLP code base Pick of th	from ReagentTable weighing more reagents' KLD and distribution etc.

Click the table icon (highlighted above) to choose from the Reagent Table

IVIAI KGI	KLD (< 1 is <info)< th=""><th>Measurement distribution</th><th>Median</th><th>Stain/fluorophor</th><th>Mean</th><th>Standard deviation</th></info)<>	Measurement distribution	Median	Stain/fluorophor	Mean	Standard deviation
CD23	1.7		-0	(Nd146)Di	0	2
Foxp3	1.65		0	(Gd158)Di	1	2
TCRgd	1.05		-0	(Dy162)Di	0	1
DNA2	0.99		111	(lr193)Di	133	85
120g8	0.97		-0	(Nd150)Di	4	22
CD11c	0.95		-0	(Nd142)Di	1	5
CD103	0.94		-0	(Yb173)Di	0	1
lgD	0.86		-0	(Pr141)Di	5	16
CD8	0.86		-0	(Eu151)Di	2	16
CD44	0.84		105	(Lu175)Di	126	115
TCRb	0.84		-0	(Tm169)Di	2	11
beadDist	0.77		30	beadDist	28	7
BC6	0.74		2	(Pd110)Di	3	4
BC3	0.73		196	(Pd105)Di	212	101
BC1	0.72		204	(Pd102)Di	220	109
BC5	0.68		4	(Pd108)Di	5	5
CD16_32	0.68		47	(Gd156)Di	70	83
CD19	0.66		-0	(Sm149)Di	2	4
NKp46	0.65		-0	(Nd145)Di	1	2
CD138	0.62		0	(Yb172)Di	1	2
BC2	0.6		161	(Pd104)Di	174	85
FceR1a	0.58		0	(Dy161)Di	1	3
CCR7	0.57		0	(Dy163)Di	1	12
Cell length	0.57		22	Cell_length	23	6
cen_lengui			1	(Gd155)Di	7	22

2.5 Select the sample (to add to the training model) by clicking **Add to the training other** related gates button

O Python's TensorFlow	Iteration limit: 1250
O MATLAB's fitcnet	% hold out for validating
Add to the training 0/9 other related gates	See in UMAF

2.6 Select BM2_cct_normalized_07_non-Neutrophils.fcs sample

	Confirm					
	Select extra gates to train on All (1/9) Sort: frequency 😧 🚮					
Ж	BM2_cct_normalized_07_non-Neutrophils.fcs BM2_cct_normalized_07_non-Neutrophils.fcs GatingTree-> View->Refresh BM2_cct_normalized_02_non-Neutrophils.fcs BM2_cct_normalized_02_non-Neutrophils.f					
	BM2_cct_normalized_08_non-Neutrophils.fcs BM2_cct_normalized_08_non-Neutrophils.fcs Manual gates 85,741/85,741 100% View->Refresh					
	BM2_cct_normalized_05_non-Neutrophils.fcs D=809 BM2_cct_normalized_05_non-Neutrophils.fcs Manual gates 84,610/84,610 100%					
	X Cancel Ok					

AutoGate indicates the number of samples added (1/9). Ensure settings for **Iteration limit** and **% hold out for validation** are as below.

Select See in UMAP check box

Train MLP on "M	lanual gates"
[^T	rain on gating sequence's
	Leaf gates only Leaf or last branch gates
Specify reagents/parameters (29	») r < ×
O Python's TensorFlow	Iteration limit: 1250
O MATLAB's fitcnet	% hold out for validating
Train on 1/9 other related gates	See in UMAP
Ca	ncel Train

2.7 Click Train

2.8 Review settings



Click OK when ready.

2.9 Save the template



2.10 Click OK to the overlap alert



2.11 MLP and UMAP Training in progress. Takes ~couple of mins.



2.12 UMAP and MLP completed.



As we have now a model, it can be applied on other samples.

3 Apply the Model

3.1 Select the manual gates of **BM2_cct_normalized_02_non_neutrophils.fcs** to apply the model.



3.2 Click Hi-D > MLP neural network under Classify/Predict

Hi-D	Edit	Search	View	Prefe	rences	Windo
Clas	sify/pr	edict				•
MLF	^o neura	l network				ЖL
EPP	' (Exha	ustive Pro	jection	Pursuit) -	∠%Ε
EPF	explai	nation				
Con	figure/	train				
MLF	^o neura	l network				ዕ ዤL
UM/	AP para	ameter ree	duction			៤ ដប
MDS	S & Hi-l	D match				N①ン
Phe	nogran	ns			trophils.	νûΡ

3.3 Ensure configuration is as below

Uncheck **Only match** (to have the gates created in the gating tree) and **By overlap** checkboxes

MLP classifying				
Classify	v via an MLP neural network			
MLP denotes " multilayer perceptron " deep learning based on feedforward fully connected neural networks				
•	 Use fitcnet Use TensorFlow (Python must be installed) 			
C 23 prior gates	3y overlap			
Classify 0/9 oth related gates	er X Cancel Ok			

Click Classify 0/9 other related gates button to select the samples that need the model applied.

Select gates for _05 and _08 samples and click OK

• • •			Cor	nfirm				
		Selec	t ext	ra gates t	o cla	assify		
	A	uli (2 /9) S	Sort:	frequer	су	\bigcirc	₽ŧ	86
		 BM2_cct_n	ormali	ized_07_nor	n-Neu	Itrophils	s.fcs	
		● ID=825	Ga Vie	tingTree-> w-> Refresh	Maı 95,0	nual ga)87/95, 100%	tes 087	
		BM2_cct_n	ormali	zed_01_nor	n-Neu	ıtrophils	s.fcs	Ш
\mathbf{X}		● ID=594	Ga Vie	tingTree-> ew-> Refresh	Mai 86,8	nual ga 864/86, 100%	tes 864	
		BM2_cct_n	ormali	zed_08_nor	n-Neu	ıtrophils	s.fcs	
		● ID=833	Ga Vie	tingTree-> w-> Refresh	Mai 85,7	nual ga /41/85, 100%	tes 741	
		 BM2_cct_n	ormali	zed_05_nor	n-Neu	ıtrophils	s.fcs	
		● ID=809	Ga Vie	tingTree->	Mai 84,6	nual ga 310/84, 100%	tes 610	
				×	Car	ncel	C)k

Confirmation that 2/9 gates are selected. Click $\ensuremath{\text{OK}}$ to run



3.4 Specify the model (filename.umap.mat)

Prompts to select the model. Specify the model that was saved earlier.



Classifying the samples using the model is quick (~ 20 secs in this example)

3.5 MLP classifications completed

Notice that the selected three samples (_02, _05 and _08) are now classified. Check the **MLP classifications gates** available for these samples. Highlighted below are for _02 and _05 samples



MLP classifications for _08 sample



4 Checking the classification using MDS

1) We have trained MLP + UMAP combined on two samples and applied that model on two separate samples. To compare between the expert (Manual) gates and MLP+UMAP gates using MultiDimensional (MDS) scaling, select both the gates at the top level or any subset of interest.



and click Hi-D > Multidimensional scaling

• •	•			Gating	Tree (AG v5	5.1.21b+):
File	Hi-D	Edit	Search	View	Preference	es Windo
🥜 태	Clas	sify/pr	edict			
(2)	MLF	o neura	l network			ЖL
BM2_	EPP EPP	(Exha explai	ustive Pro nation	jection I	Pursuit)	∕гже
8	Con	figure/	train			
0	MLF	o neura	l network			<mark>ଫ</mark> <mark>೫</mark> ∟
~	UMA	AP para	ameter red	duction		<mark>ଫ</mark> װ
	MDS	5 & Hi-	D match			τôΜ
	Phe	nogran	ns			τûΡ
	Visu	alize				
	UMA	AP para	ameter red	duction		てむし
	MDS	S (mult	idimensio	nal scali	ng)	М
	Phe MDS	nogran S & phe	n(s) enogram(s	s)		Ϋ́Ρ
	Hea	t map				ΥE

2) Brings up the MDS/Hi-D match window with Manual gates on the left and MLP gates on the right



3) For a detailed view, click the table icon from the tool bar



Brings up the QF match summary table

ank ^{^1}	#	Name	Mat- ches	Similarity (QFMatch)	Overlap (F-measure)	Plot	2 Freq. 3	# of events	Gate ID	Orphans	Parent orphans	Non-EP orphans
1	• 23	B cells Frac. F	2	99.9%	98.8%	1	16.7%	14,671	755	N/A	N/A	N/A
1	• 59	B cells Frac. F	2	99.9%	98.8%	2	17%	14,903	2050	N/A	N/A	N/A
2	• 1	Eosinophils	2	99.9%	98.7%	1	5.7%	4,991	681	N/A	N/A	N/A
2	• 25	Eosinophils	2	99.9%	98.7%	2	5.7%	4,998	2016	N/A	N/A	N/A
3	• 3	pDCs	2	99.8%	98.5%	1	3%	2,677	689	N/A	N/A	N/A
3	• 29	pDCs	2	99.8%	98.5%	2	3.1%	2,682	2020	N/A	N/A	N/A
4	• 22	B cells Frac. E	2	99.7%	97.1%	1	7.3%	6,420	754	N/A	N/A	N/A
4	• 58	B cells Frac. E	2	99.7%	97.1%	2	7.5%	6,619	2049	N/A	N/A	N/A
5	• 8	CD8 T cells	2	99.6%	98.1%	1	1.9%	1,650	710	N/A	N/A	N/A
5	• 37	CD8 T cells	2	99.6%	98.1%	2	1.9%	1,650	2028	N/A	N/A	N/A
6	• 17	Classical Monocytes	2	99.6%	98.1%	1	14%	12,245	742	N/A	N/A	N/A
6	<mark>-</mark> 51	Classical Monocytes	2	99.6%	98.1%	2	13.9%	12,231	2042	N/A	N/A	N/A
7	• 21	B cells Frac. D	2	99.5%	95.5%	1	2.1%	1,856	752	N/A	N/A	N/A
7	• 57	B cells Frac. D	2	99.5%	95.5%	2	2.1%	1,849	2048	N/A	N/A	N/A
8	• 18	Intermediate Monocytes	2	99.5%	95.6%	1	4.8%	4,221	744	N/A	N/A	N/A
8	<mark>-</mark> 52	Intermediate Monocytes	2	99.5%	95.6%	2	4.9%	4,282	2043	N/A	N/A	N/A
9	• 19	Non classical monocytes	2	99.3%	95%	1	0.9%	802	745	N/A	N/A	N/A
9	• 53	Non classical monocytes	2	99.3%	95%	2	0.9%	826	2044	N/A	N/A	N/A
10	• 7	CD4 T cells	2	99.3%	95%	1	0.9%	816	708	N/A	N/A	N/A
10	• 36	CD4 T cells	2	99.3%	95%	2	0.9%	813	2027	N/A	N/A	N/A
11	• 11	MEP	2	98.9%	92.1%	1	1.1%	972	727	N/A	N/A	N/A
11	<mark>.</mark> 43	MEP	2	98.9%	92.1%	2	1.2%	1,029	2034	N/A	N/A	N/A
12	• 5	Basophils	2	98.8%	93%	1	0.5%	423	701	N/A	N/A	N/A
12	• 32	Basophils	2	98.8%	93%	2	0.5%	431	2023	N/A	N/A	N/A
13	• 20	B cells Frac A-C	2	98.7%	88.2%	1	1.5%	1,356	748	N/A	N/A	N/A
13	• 55	B cells Frac A-C	2	98.7%	88.2%	2	1.5%	1,296	2046	N/A	N/A	N/A
14	• 10	Macrophages	2	98.5%	90.9%	1	0.5%	477	717	N/A	N/A	N/A
14	• 41	Macrophages	2	98.5%	90.9%	2	0.5%	471	2032	N/A	N/A	N/A
15	• 2	Plasma Cells	2	98.1%	94.6%	1	0.2%	170	685	N/A	N/A	N/A
15	. 27	Plasma Cells	2	98.1%	94.6%	2	0.2%	181	2018	N/A	N/A	N/A

4) To get the mass+distance similarity score (using QFMatch) and F-measure overlap, click on the below buttons in the toolbar



Shows the accuracy level.



4.1 Investigating classification quality in further detail

The median and mean summaries of the test set's similarity and overlap with a training set classification is not likely adequate to form conclusions about a particular configuration of MLP/UMAP classifiers. Thus, additional views are provided.

The toolbar button on the Hi-D match table with label +- allows you to create the other supplementary visualizations of the classification.

MATCH sets: training=all 23 matched!, test=23/36 ^(13 NEW!)
 False positive/negative charts
 Prediction detail table
 Confusion chart
 Import false +/- into AutoGate GatingTree

4.2 Prediction Adjudicator

Click on the Prediction detail table from the above menu to bring up the Prediction Adjudicator table. This table aids the user in adjudicating whether for each particular class the false negatives (classifications made only by the biologist () have greater mass+distance similarity to the whole predicted class than false positives (classifications made only by MLP/UST).

$\bullet \bullet \bullet$

PredictionAdjudicator 36 X 66 subsets

🍓 🌍 🗐 🔶 💟 4 📰 💿 1 📰 👰 False + MORE similar in 2/15 cases

Subset (class) name	Similarity to	# of events	Freq.	#	Overlap (F-measure)	•	Subset type	2 Subset ID
CD8 T cells false +	77%	32	0%	65	0%	•	false +	2028.2
CD8 T cells false -	83.2%	32	0%	66	3.8%	•	false -	2028.3
CD3- CD49b (non specific) false -	100%	6	0%	67	100%	•	false -	2029.3
mDCs ^{true +}	96.4%	154	0.2%	68	91.4%	•	true +	2030.1
mDCs false +	82.3%	20	0%	69	0%	•	false +	2030.2
mDCs ^{false -}	80.8%	29	0%	70	27.4%	•	false -	2030.3
non-mDCs (non specific) ^{false -}	100%	7,540	9.9%	71	100%	•	false -	2031.3
Macrophages ^{true +}	98.9%	431	0.6%	72	95.6%	•	true +	2032.1
Macrophages false +	80.5%	46	0.1%	73	0%	•	false +	2032.2
Macrophages false -	88.5%	40	0.1%	74	15.7%	•	false -	2032.3
Lin-cKit+ Sca1- (non specific) ^{false -}	100%	430	0.6%	75	100%	•	false -	2033.3
MEP true +	98.5%	921	1.2%	76	94.5%	•	true +	2034.1
MEP false +	80.3%	51	0.1%	77	0%	•	false +	2034.2
MEP false -	86.9%	108	0.1%	78	19%	•	false -	2034.3
GMP true +	97.6%	1,338	1.8%	79	92.6%	•	true +	2035.1
GMP false +	83.2%	76	0.1%	80	0%	•	false +	2035.2
GMP false -	84.8%	214	0.3%	81	24.2%	•	false -	2035.3
CMP false -	100%	402	0.5%	82	100%	•	false -	2036.3
HSC ^{false -}	100%	21	0%	83	100%	•	false -	2037.3
MPP false -	100%	301	0.4%	84	100%	•	false -	2038.3
Lin cKit- Sca1+ (non specific) ^{false -}	100%	320	0.4%	85	100%	•	false -	2039.3
CLP false -	100%	163	0.2%	86	100%	•	false -	2040.3
Monocytes (non specific) ^{false -}	100%	1,777	2.3%	87	100%	•	false -	2041.3
Classical Monocytes false -	100%	12,231	16.1%	88	100%	•	false -	2042.3
Intermediate Monocytes false -	100%	4,282	5.6%	89	100%	•	false -	2043.3
Non classical monocytes false -	100%	826	1.1%	90	100%	•	false -	2044.3

4.3 Confusion chart

Chart shows the agreement between two classifications in the diagonal boxes and the other boxes are the ones with false+/-



True Class = Biologist classification, Predicted Class = MLP classification

5 Visualize gates in Parameter Reduction / Plot Editor

1) To view a MLP gate, double click the gate. Say **Ok** to open the gate in **Parameter Reduction** window



The above prompt is only shown if **Preferences** is set to **Ask if gate is MLP+UMAP** (as shown below)





2) Opens the gate in Parameter Reduction window

Select the subsets to have it reflected in the Parameter Reduction window, as below.



3) To view the gate in Plot Editor window, select the gate and click on the MLP icon highlighted below



4) Opens in Plot Editor window

🔴 😑 🌒 Plote	ditor (AG v5.1.21b+): BM2_cct_normalized_0	02_non
🆫 🖬 🖪 🖕	🔲 🕅 🕂 👄 Recent axes 😒	1
🥔 🗄 🖌 🏷	🍥 0/1 clusters 😌 🌆 M 🚠 🔷 🔻	(v) »
Plot 3 has 14,903 ever	nts 14k (98.8%) events are clustered	Preferences
104-	• B cells Frac. F ^(14,903/87,772 17%)	L
10 ³		
€ 10 ²		
> 10 ¹		
0 -	0 10 ¹ 10 ² 10 ³	104
10 - C	FceR1a:(Dy161)Di ▼	

5) Here on, you can continue with gating operations as before (e.g, to change the X/Y parameters, click on the Reagent table (highlighted above) and set in the table.

