Applications of computation and artificial intelligence in clinical cytometry

Paul D. Simonson, M.D., Ph.D.
Weill Cornell Medicine

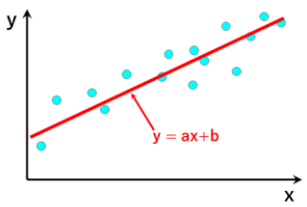
ACS 2024 Conference
October 22, 2024, 9:30 a.m.

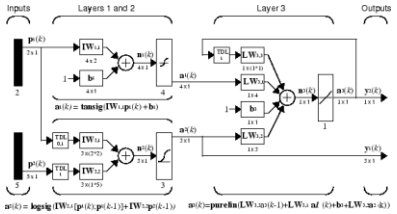
Outline

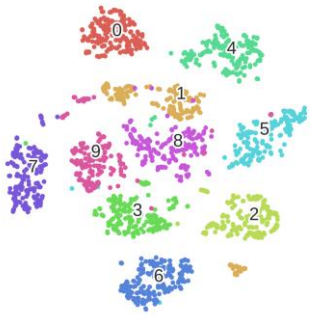
- What is machine learning?
- How can machine learning help in flow cytometry?
- Basics of supervised machine learning
- Example application: Diagnosis of classic Hodgkin lymphoma
- Comparing machine learning classifiers
- Example application: Predicting need for add-on panel
- Challenges
- Summary

What is machine learning?

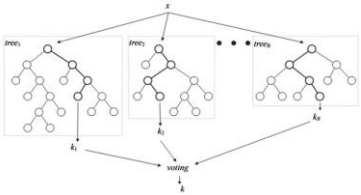
- Statistics
- Identifying patterns
- Learns what we teach it
- "Models"

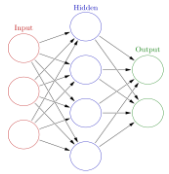






- Two general categories:
- Unsupervised
  - Clustering algorithms (e.g. Flow SOM)
  - Dimensionality reduction (e.g., UMAP)
- Supervised
  - Random forests
  - Support vector machines
  - Deep learning





What machine learning is not...

- Replacement for pathologists
- Sentient robots
- The end of humanity...yet?

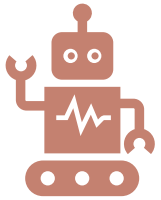
How can machine learning help with flow cytometry?

Better accuracy

Less subjectivity

Alerts for possible missed diagnoses

Highlight cells of interest



Faster workflow

Automate ordering of additional antibodies

Alerts for high priority diagnoses

Simplify increasingly complex analyses

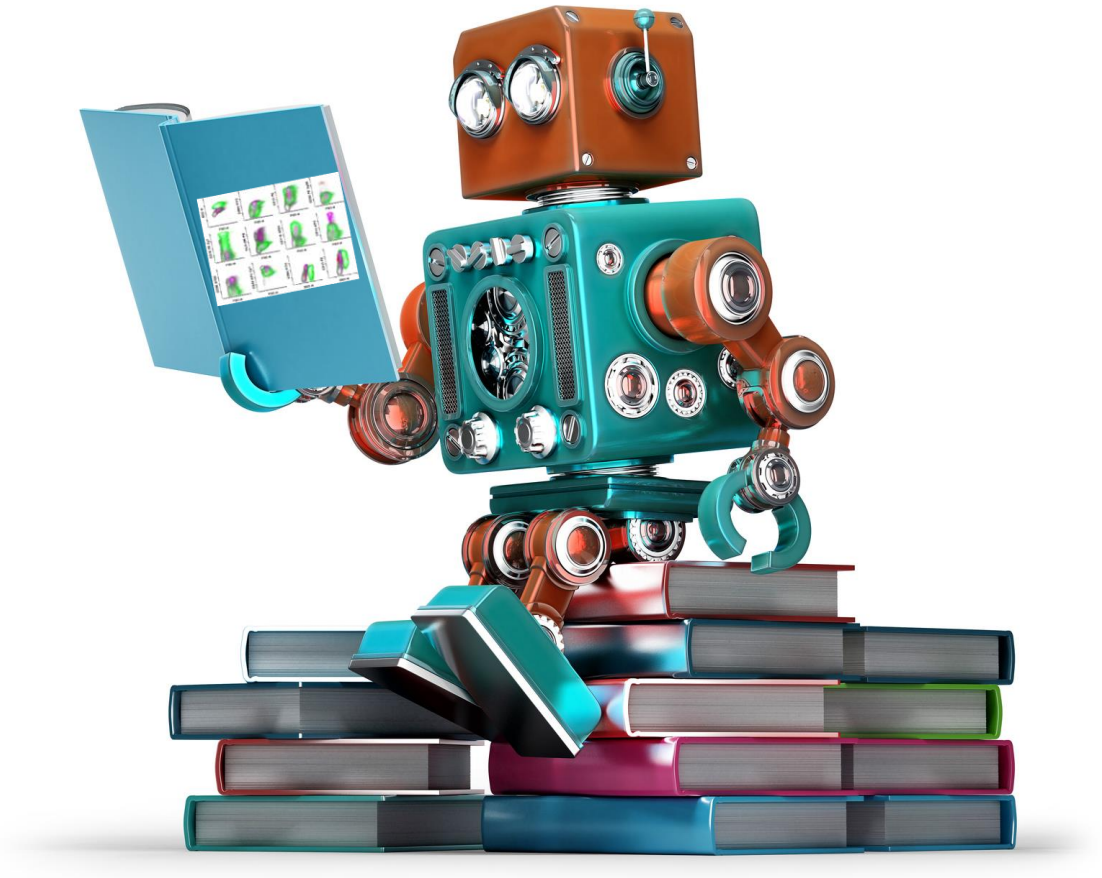
Pre-populate reports

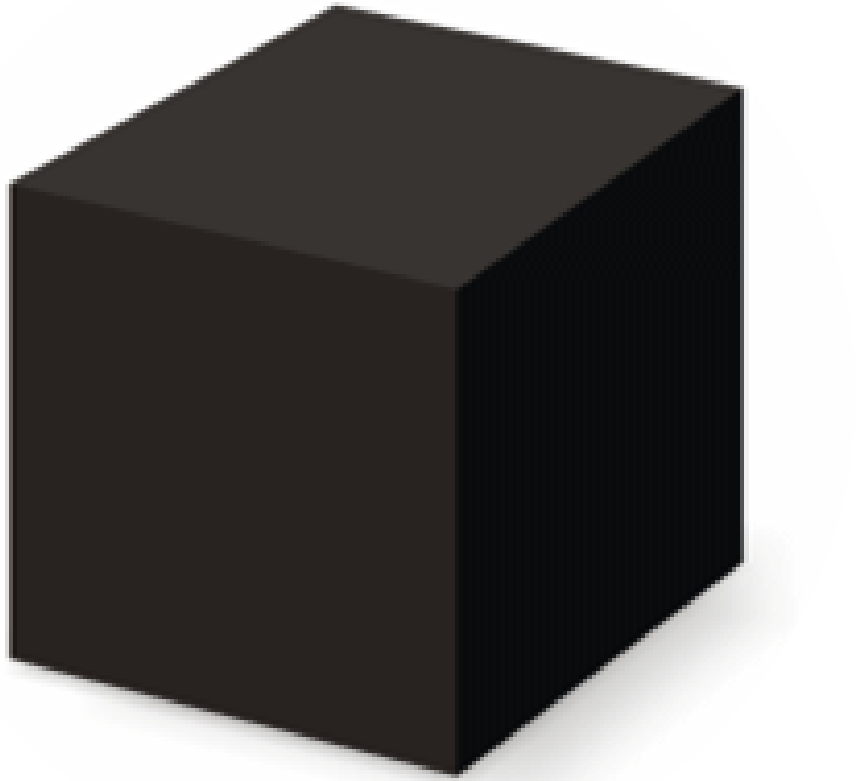
Reduced workload

Shorter work week???

Opportunities for machine learning in flow cytometry: many!

- Workflow
  - Initial specimen processing (orders, prior history, initial antibody panel choices)
  - Flow the specimen
  - Data review by technologists
    - Additional antibody panels?
    - Draft report
  - Review by hematopathologist
    - Additional antibody panels?
    - Final report
    - If new diagnosis, review by second hematopathologist
    - Improvements in accuracy?
    - Improvements in clinical outcomes?

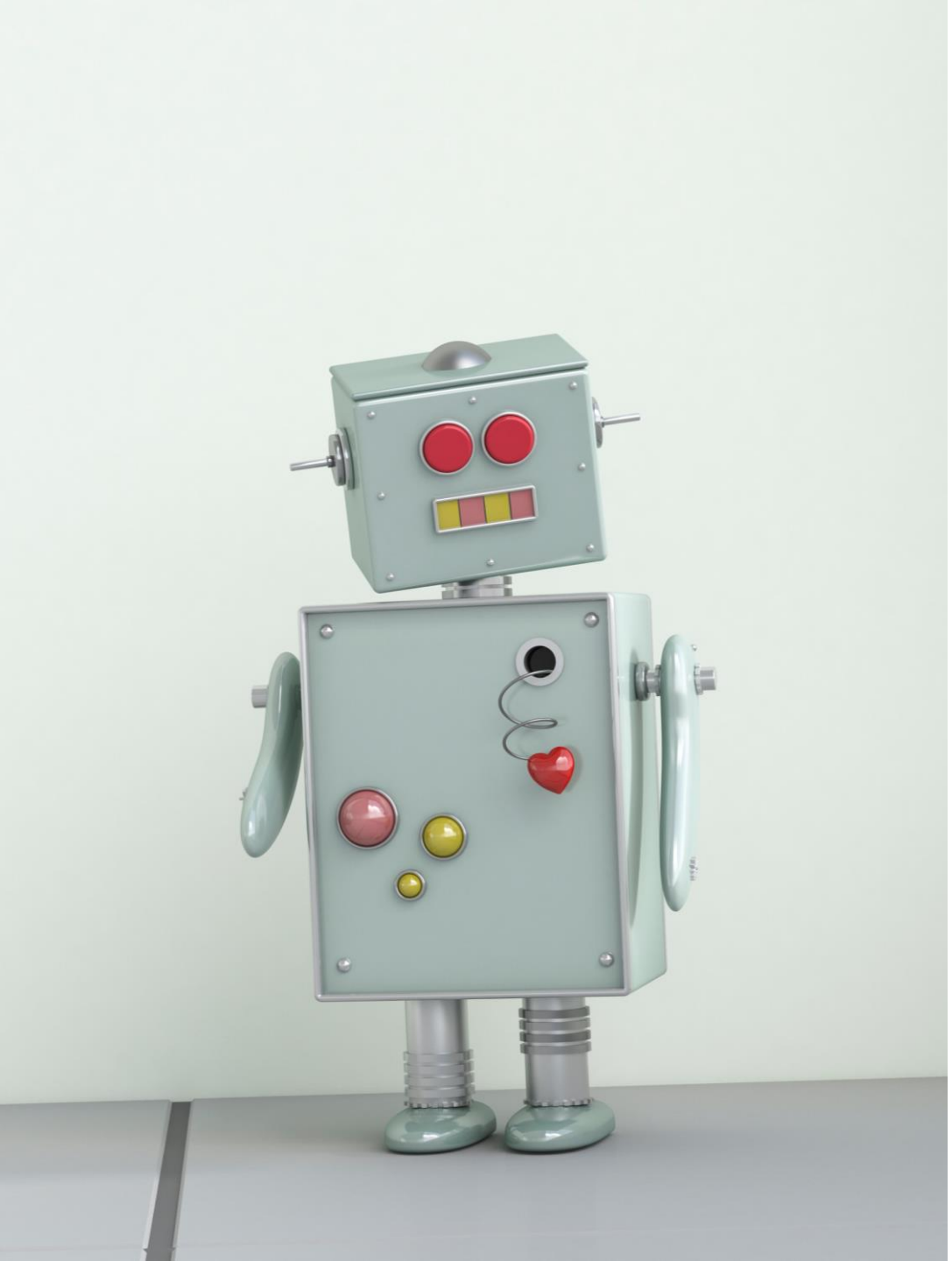




Machine learning is often a black box

- The inner workings of the machine are complex.
- Nobody likes a black box.
  - Physicians
  - Machine learning practitioners
  - Regulators
- How do we look inside?
  - Multiple methods (SHAP, LIME, etc.); can depend on the machine learning algorithm used.
  - Subject of research by many groups.
  - Methods are getting better all the time.

https://towardsdatascience.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0

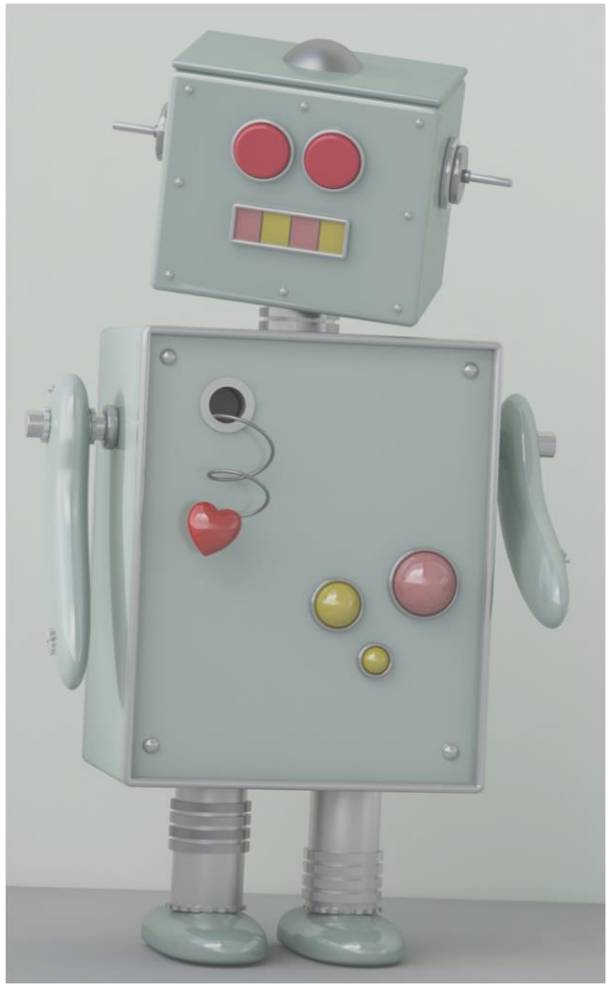


FlowBot: “dumb AI” can actually do a lot of work!

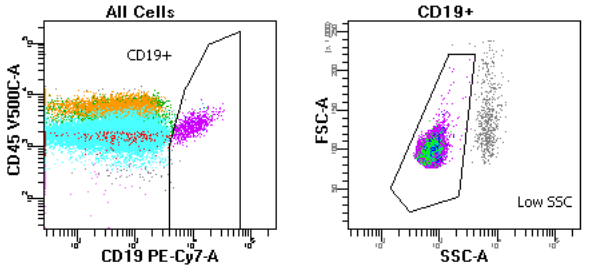
- Machine learning development takes a lot of development
- Is there a “poor man’s AI” we can use in the meantime?

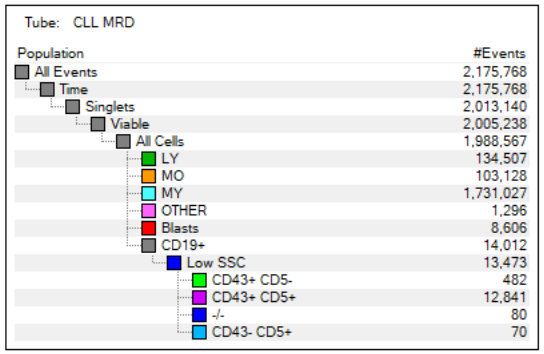
Rules-based automated report generation

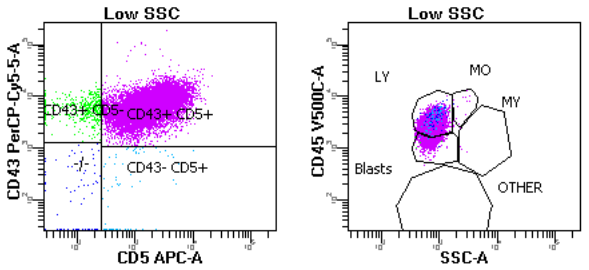
- Approach:
  - Computer running “Flowbot” python script checks file server every 10 minutes for newly completed flow analysis files
  - Antibody panels are identified, as well as gated populations’ numbers
  - Text file with autogenerated report is written, based on a series of if-then programming statements
  - Running list of day’s cases with possible neoplasm is updated
  - Text file can be copied, modified, or ignored by pathologist
- Advantages
  - Algorithm is fully understood
  - Completely customizable
- Disadvantages
  - Not very fancy
  - Does not learn

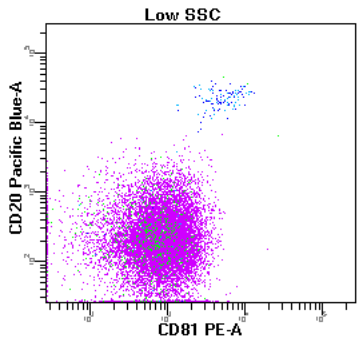


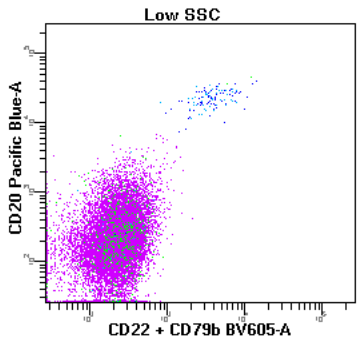
Example: CLL MRD

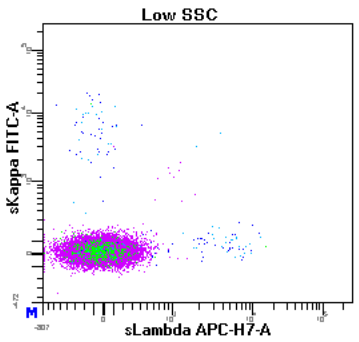












INTERPRETATION

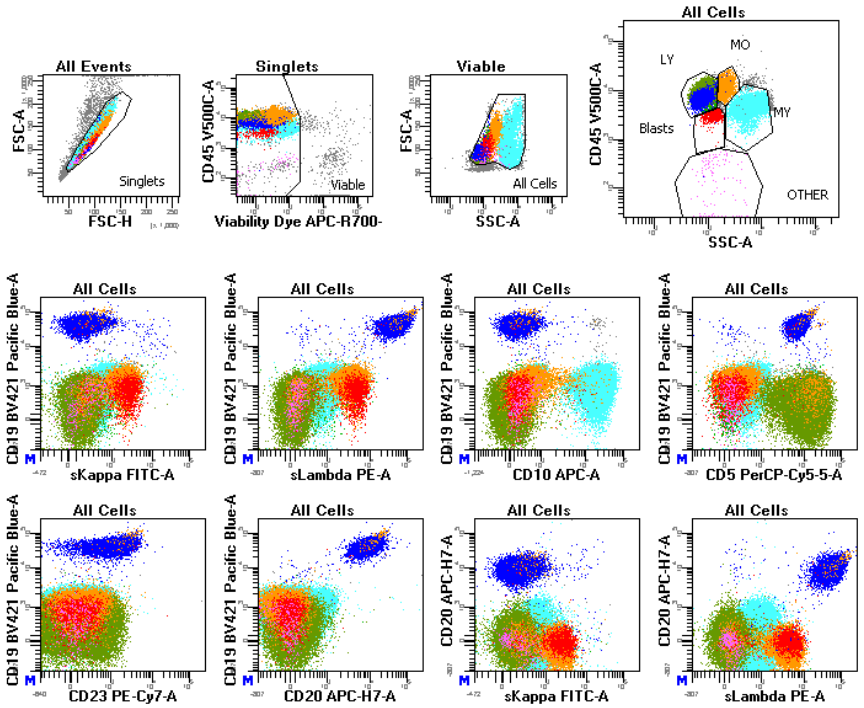
Cytospin: Morphologic evaluation of a Diff-Quik-stained cytospin demonstrates peripheral blood elements.

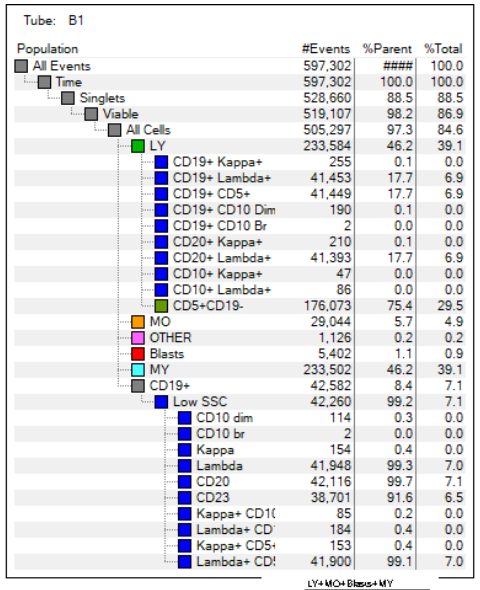
Flow cytometry: Analysis was performed using the antibodies listed above. CD19+, CD5+, CD43+, ___-restricted B cells are 0.6% of viable cells. Lymphocytes, monocytes, and granulocytes are approximately 6.8%, 5.2%, and 87% of viable cells, respectively.

The flow cytometry findings are consistent with persistent chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL).

A minimum of one million viable cells were analyzed for MRD assessment to achieve an analytical sensitivity of 0.01%.

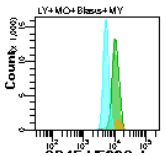
Example: B cell neoplasm





Tube: B1

Population	#Events	%Parent	%Total
All Events	597,302	####	100.0
Time	597,302	100.0	100.0
Singlets	528,660	88.5	88.5
Viable	519,107	98.2	86.9
All Cells	505,297	97.3	84.6
LY	233,584	46.2	39.1
CD19+ Kappa+	255	0.1	0.0
CD19+ Lambda+	41,453	17.7	6.9
CD19+ CD5+	41,449	17.7	6.9
CD19+ CD10 Dim	190	0.1	0.0
CD19+ CD10 Br	2	0.0	0.0
CD20+ Kappa+	210	0.1	0.0
CD20+ Lambda+	41,393	17.7	6.9
CD10+ Kappa+	47	0.0	0.0
CD10+ Lambda+	86	0.0	0.0
CD5+CD19-	176,073	75.4	29.5
MO	29,044	5.7	4.9
OTHER	1,126	0.2	0.2
Blasts	5,402	1.1	0.9
MY	233,502	46.2	39.1
CD19+	42,582	8.4	7.1
Low SSC	42,260	99.2	7.1
CD10 dim	114	0.3	0.0
CD10 br	2	0.0	0.0
Kappa	154	0.4	0.0
Lambda	41,948	99.3	7.0
CD20	42,116	99.7	7.1
CD23	38,701	91.6	6.5
Kappa+ CD1(	85	0.2	0.0
Lambda+ CD	184	0.4	0.0
Kappa+ CD5	153	0.4	0.0
Lambda+ CD5	41,900	99.1	7.0



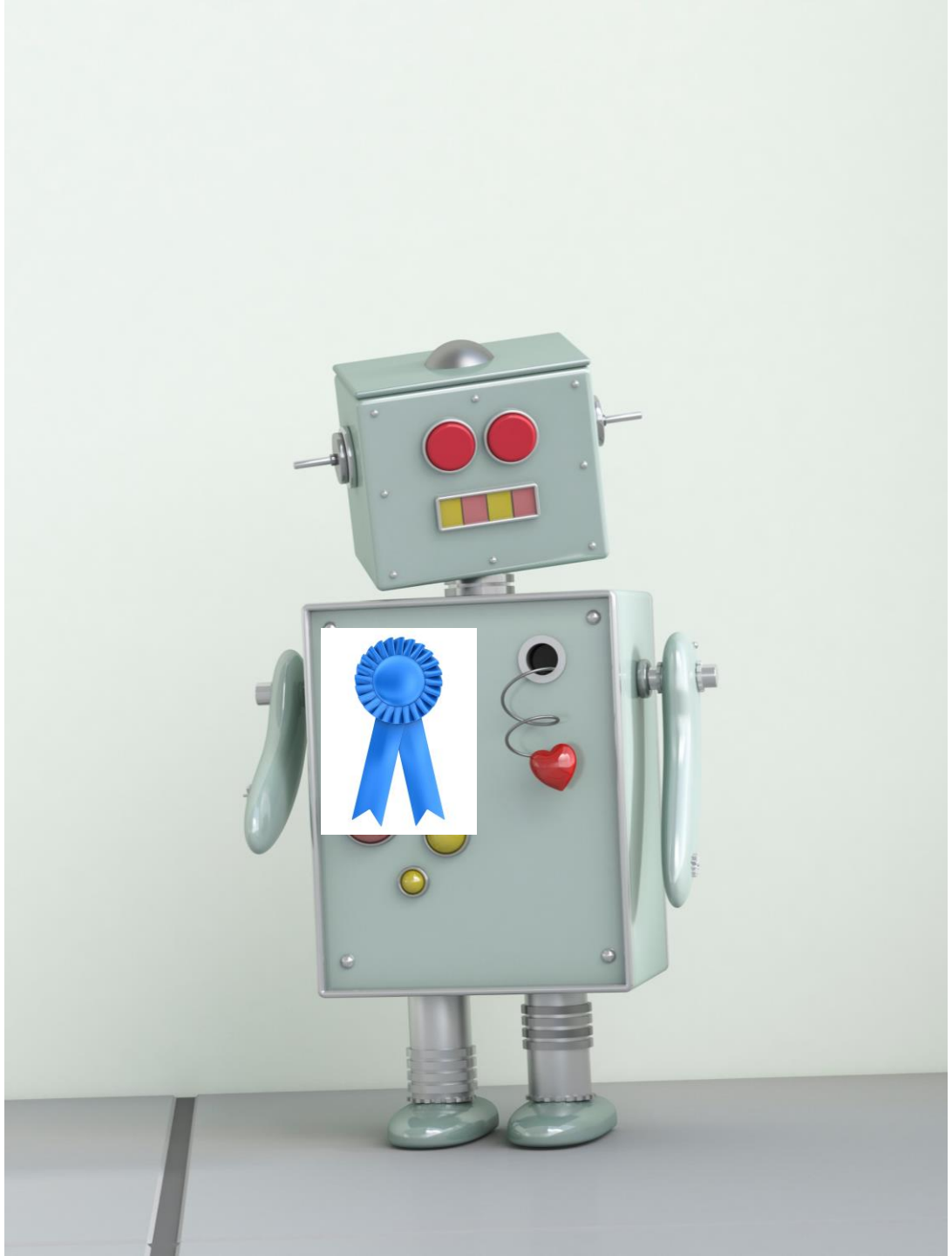
Peripheral blood sample. Panels run: B1, B2, T1, PCD1.

INTERPRETATION

Cytospin: Morphologic evaluation of a Diff-Quik-stained cytospin demonstrates peripheral blood elements.

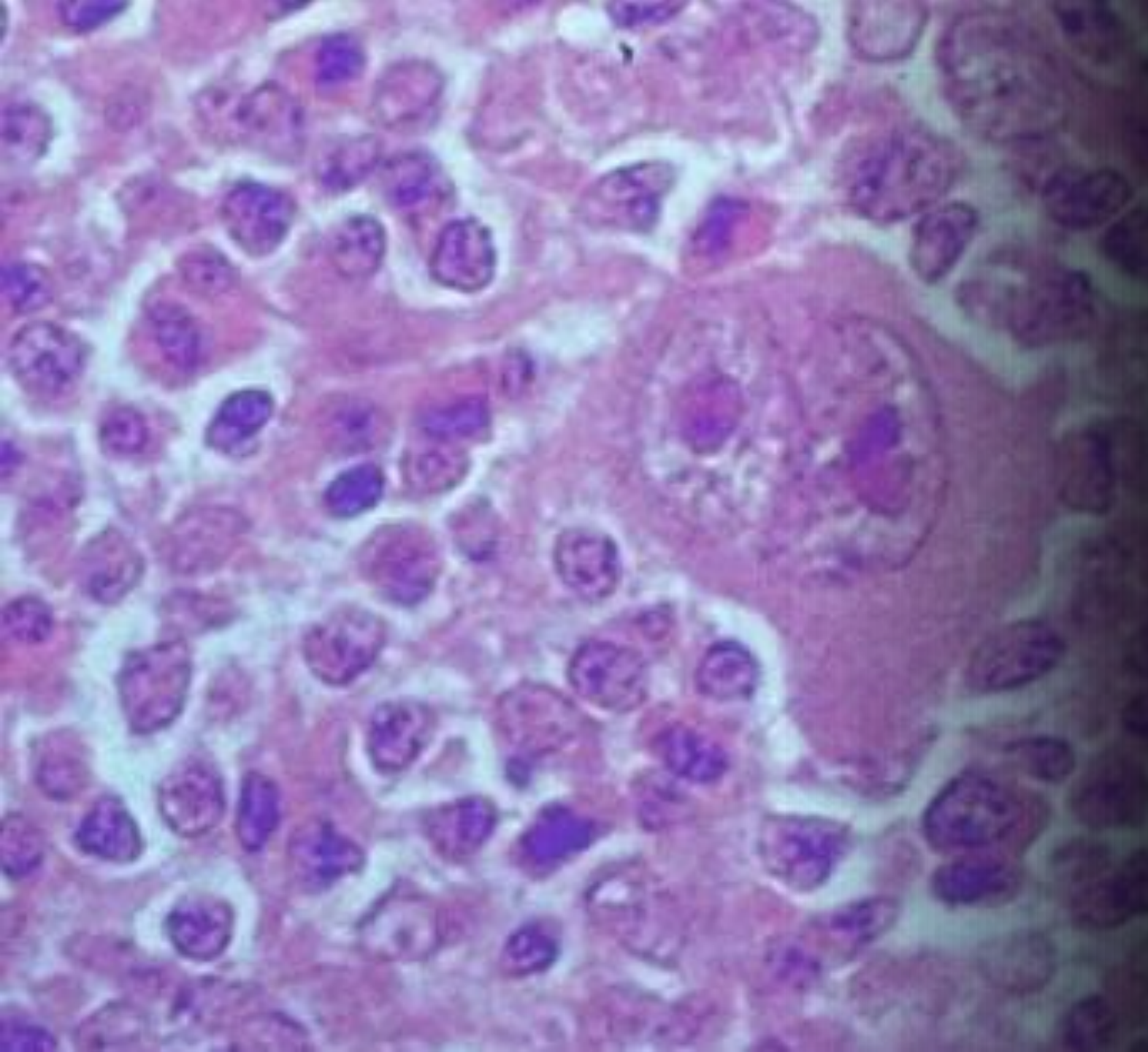
Flow cytometry: Analysis was performed using the antibodies listed above. There is an abnormal, lambda-restricted B cell population (8.2% of viable cells, 18% of lymphocytes, 99% of B cells) that co-expresses ____ and is negative for ____ expression. T cells show no loss of pan-T-cell antigens. The CD4:CD8 T cell ratio is 1.2:1. Plasma cells are too few in number for reliable evaluation for light chain restriction. Lymphocytes, monocytes, and granulocytes are approximately 46%, 5.7%, and 46% of viable cells, respectively.

The flow cytometry findings are compatible with a B cell neoplasm. There is no diagnostic evidence of a T cell or plasma cell neoplasm by flow cytometry.



General steps in applying supervised machine learning

1. Identify and annotate training data
2. Select a classifier algorithm (random forest, CNN, GMM, etc.)
3. Train the classifier
4. Apply the classifier to a testing/validation data set
5. Implement prospectively

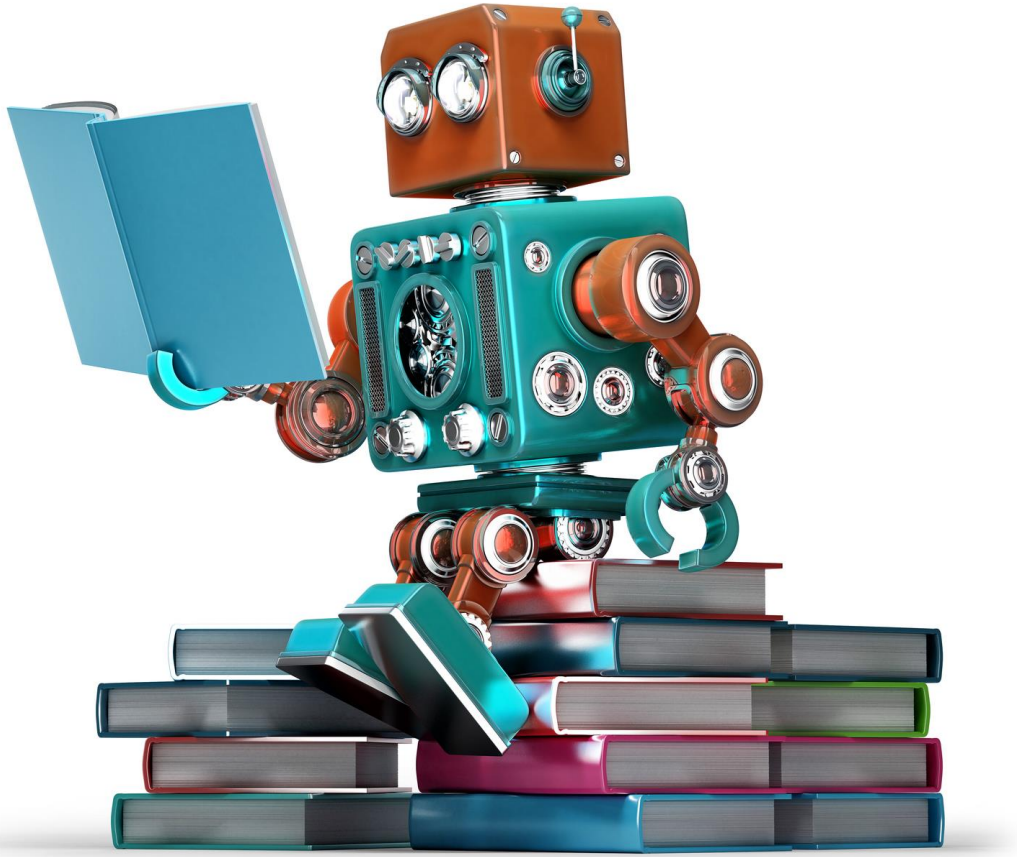


Example application: Detecting classic Hodgkin lymphoma

Objectives and approach

- Develop a machine learning algorithm that can predict the diagnosis based on flow.
- Design with the intention of applying interpretability algorithms.

- Create a series of two-dimensional histograms as inputs to the algorithm.
- Use an ensemble of convolutional neural networks (CNNs) for the classifier.

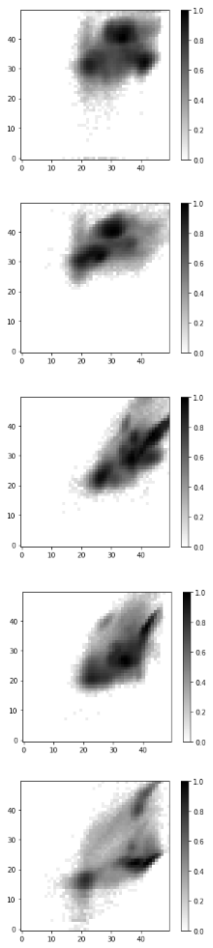


Data

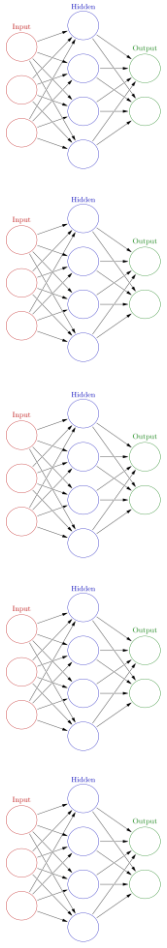
- 1222 patient cases (2010 to mid-2019)
  - 321 positive or suspicious for cHL
  - 921 negative
    - 180 cases consistent with or suspicious for a different neoplasm.
- 80% used for training, 20% for testing

The model used: “EnsembleCNN”

Raw Data File (FCS)



Histograms



CNNs

75%

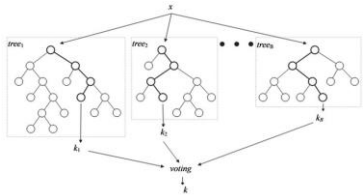
71%

40%

65%

85%

Intermediate Predictions



Random Forest Classifier

88%

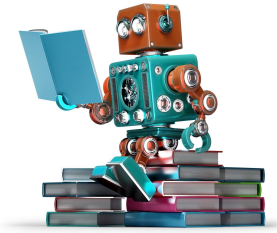
Final Prediction

Benefits:

- Intuitive (similar to human approach)
- Flexible
- Amenable to explainability tools

Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

Step 1: Train the neural networks (CNNs)



Raw Data File

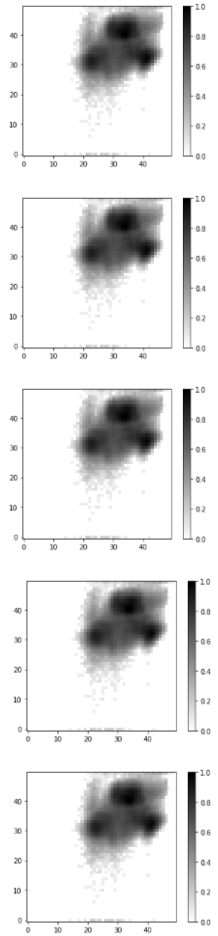
Raw Data File

Raw Data File

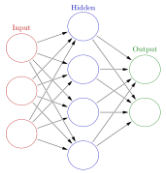
Raw Data File

Raw Data File

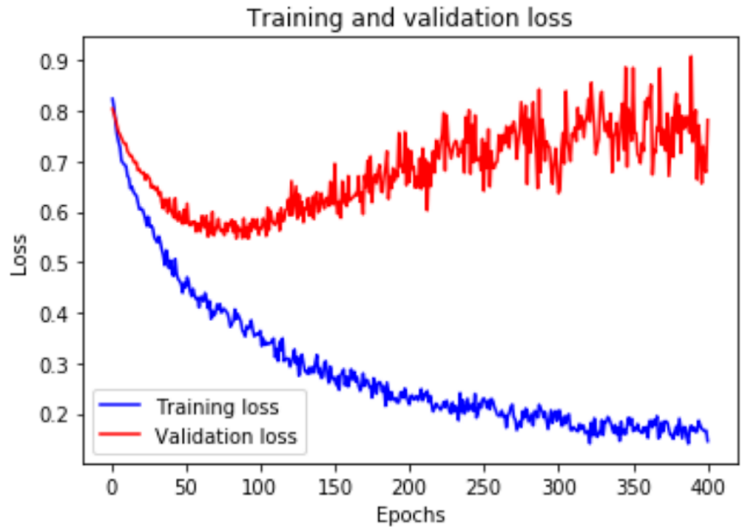
Training data

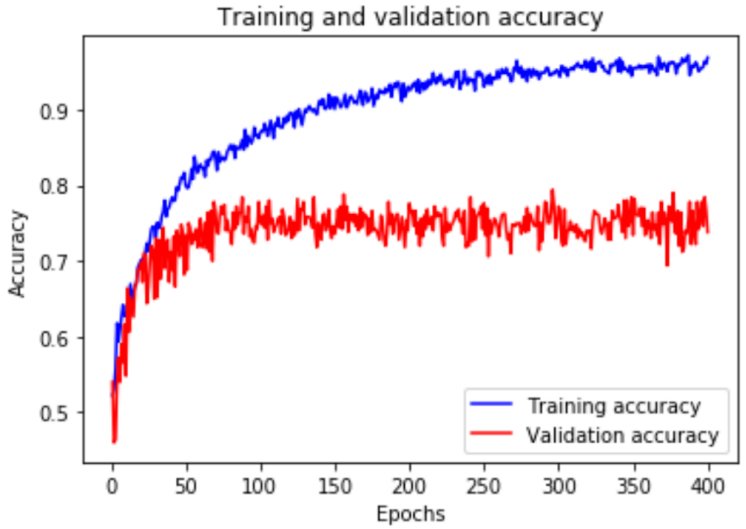


Histograms (e.g., CD20 vs. CD5)



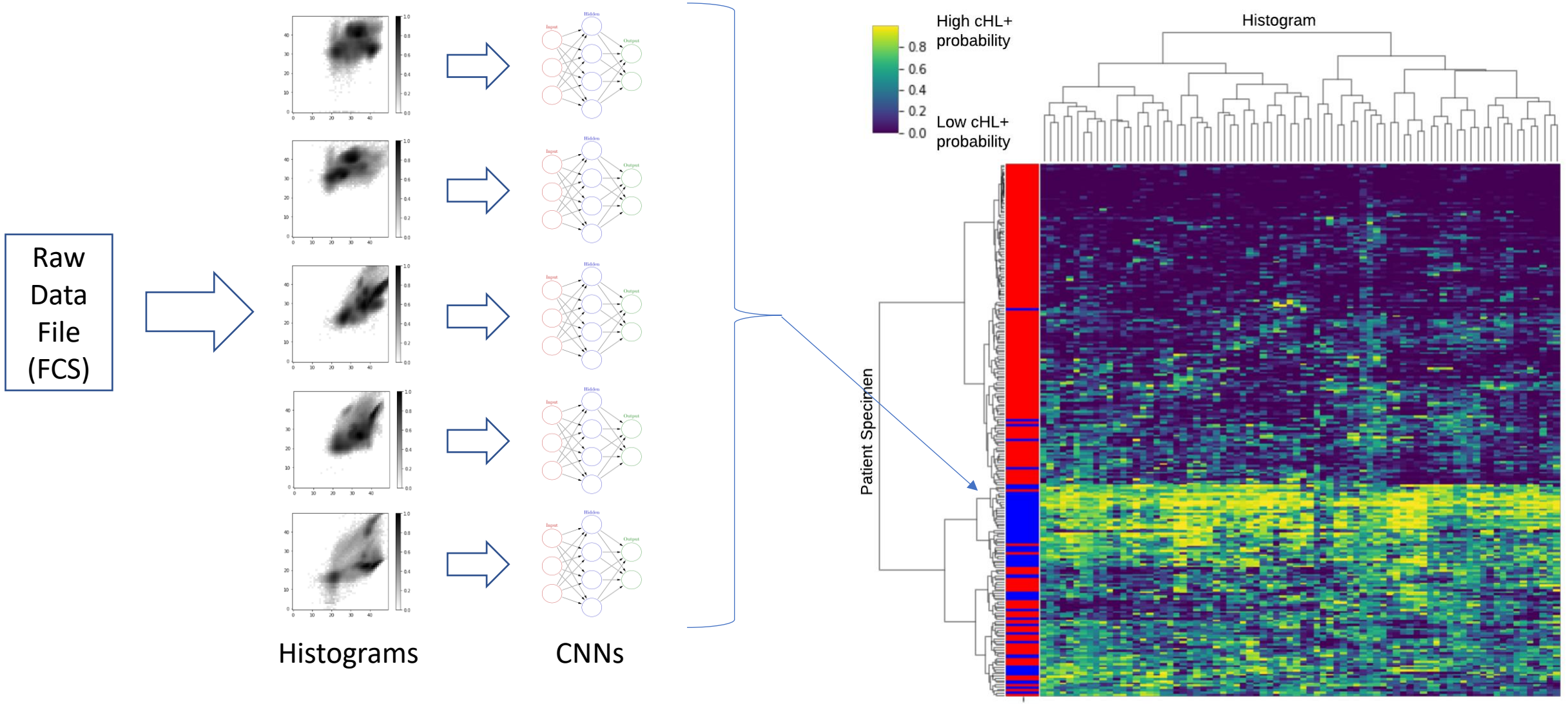
CNN





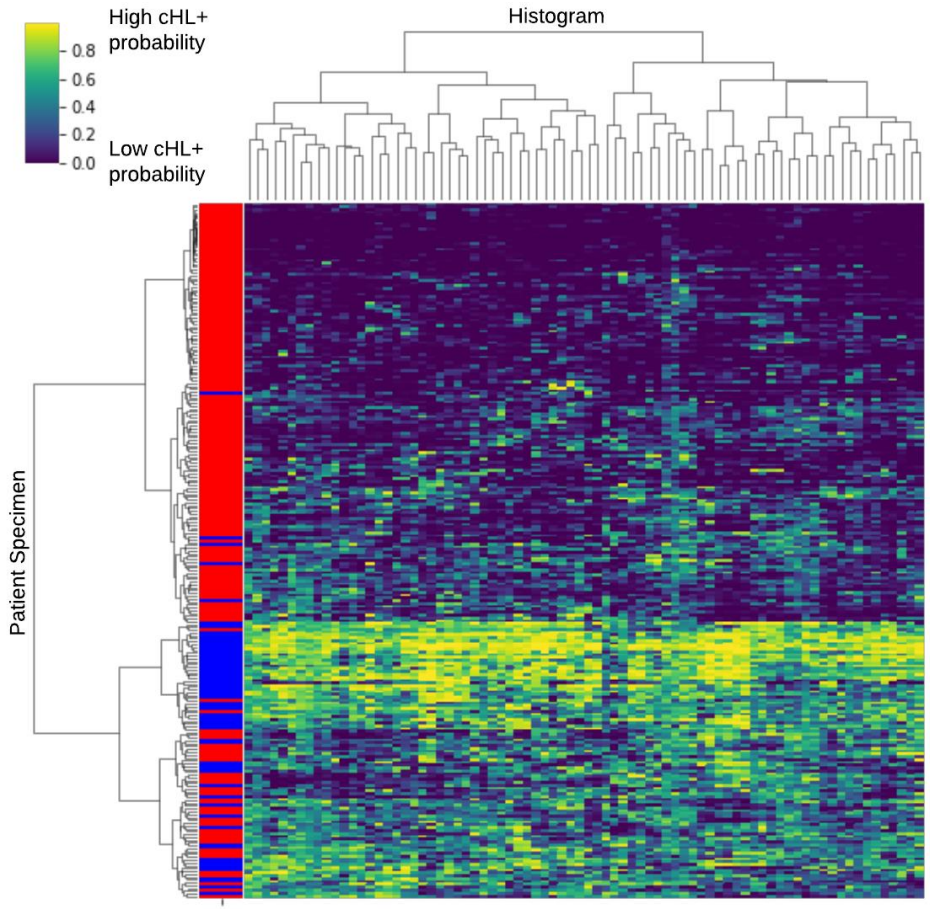
Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

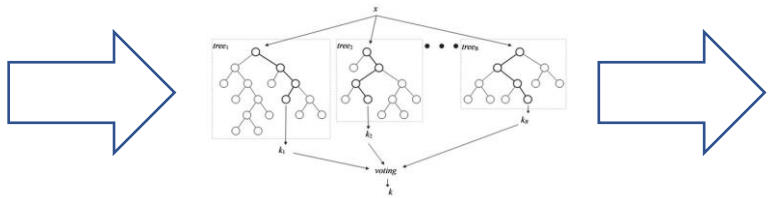
Step 2: Use the CNNs to predict results



Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

Step 3: Use the results to train the random forest classifier





Random Forest Classifier

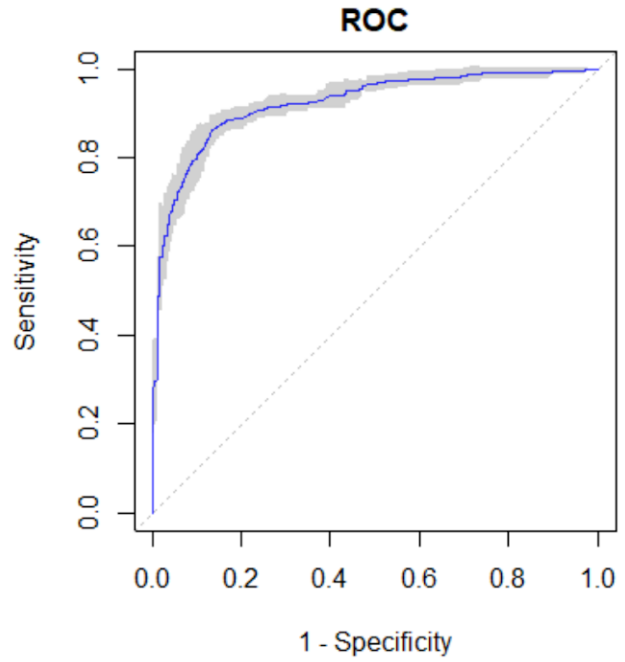
Final Predictions

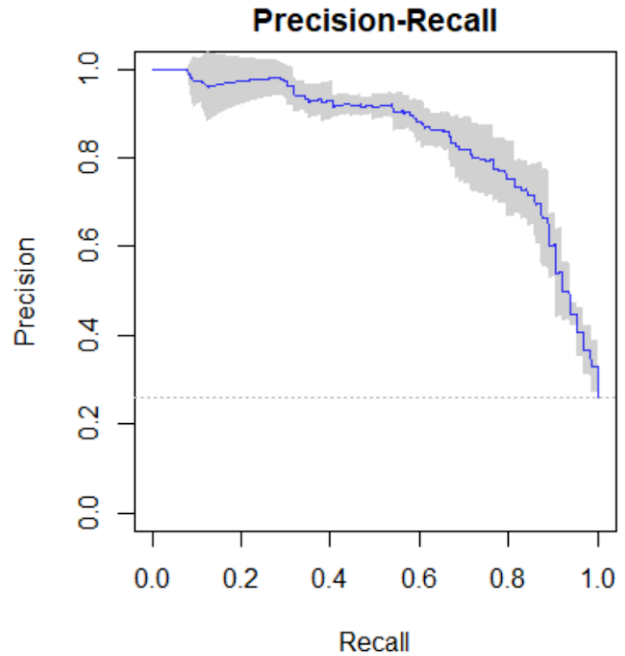
Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

Final Results

- 89% accuracy
- AUC = 0.93
- Precision = 83%
- Recall (sensitivity) = 69%
- F1 score = 75%

CNN ensemble classifier	Predicted -	Predicted +
True -	174	9
True +	20	42

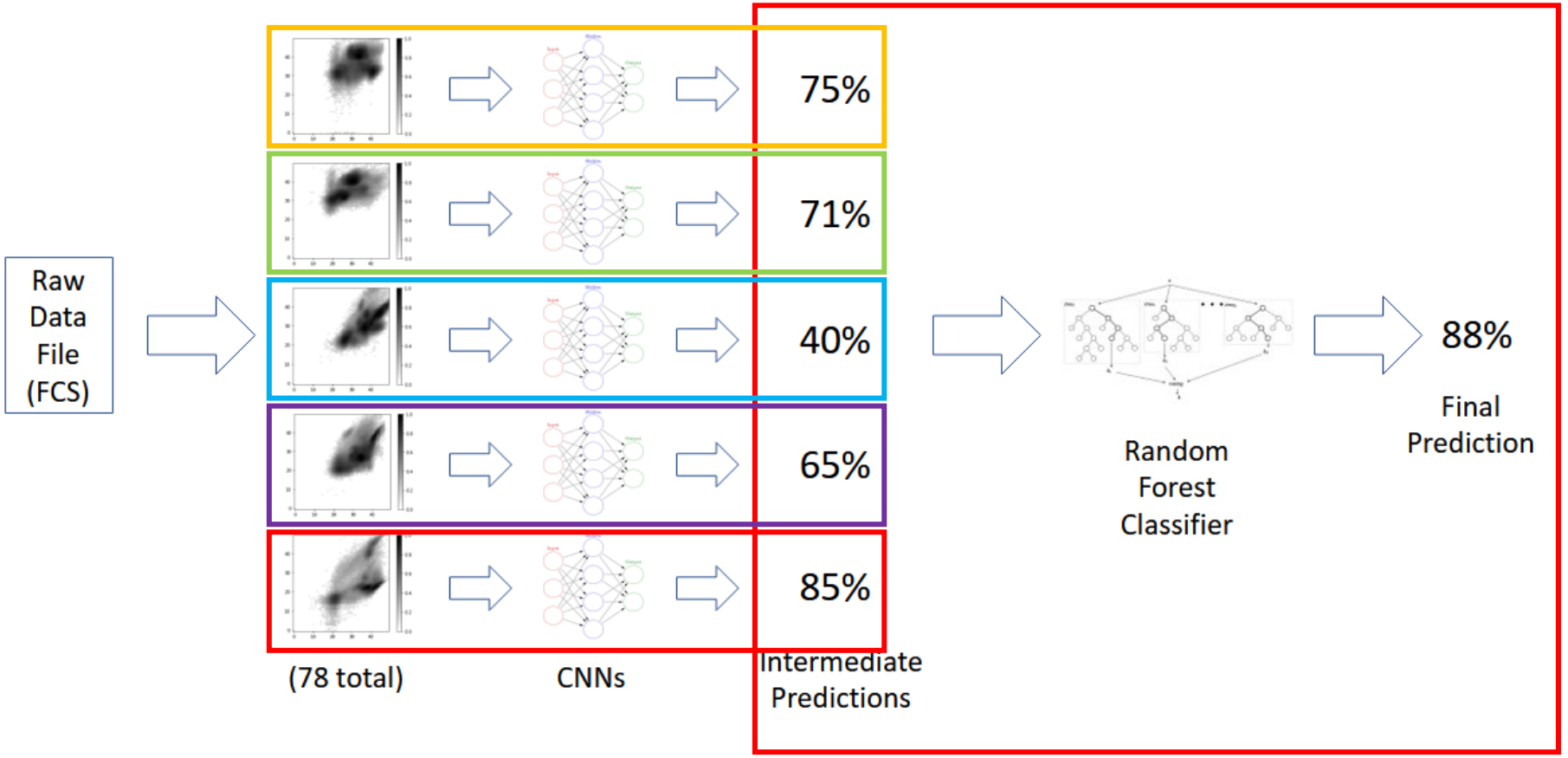




Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

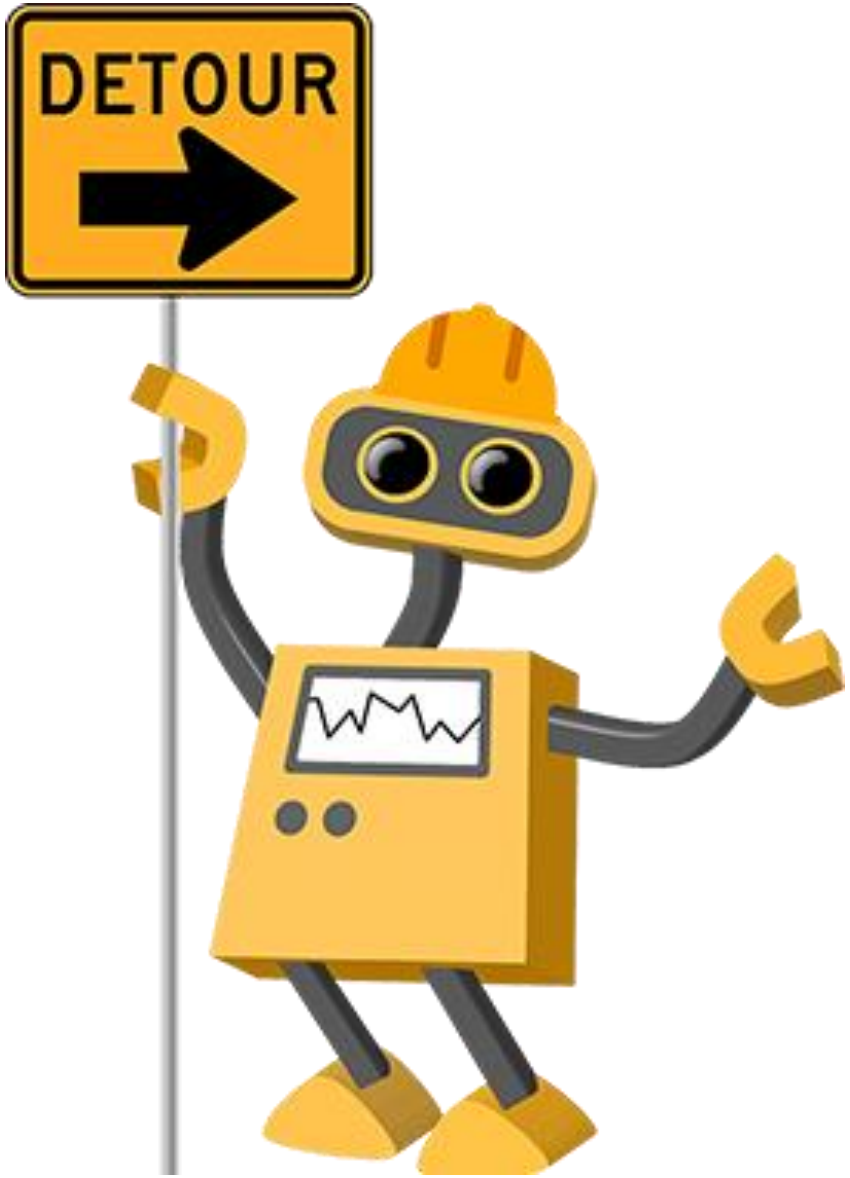
That's great, but how do you explain it?

- Break the process up into parts
- Metric to use: SHAP values



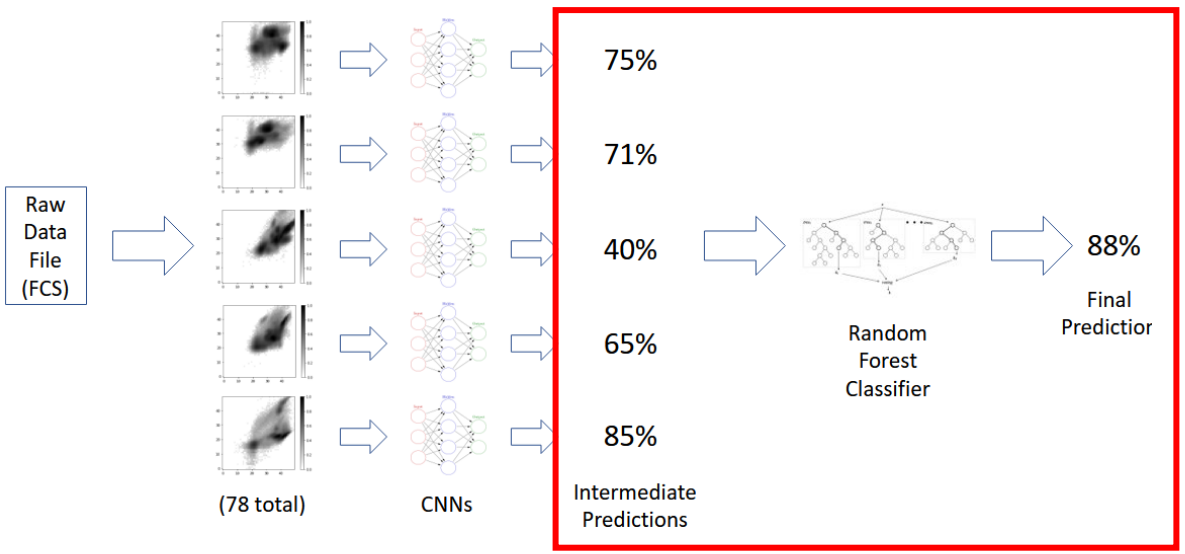
Detour: What are SHAP values?

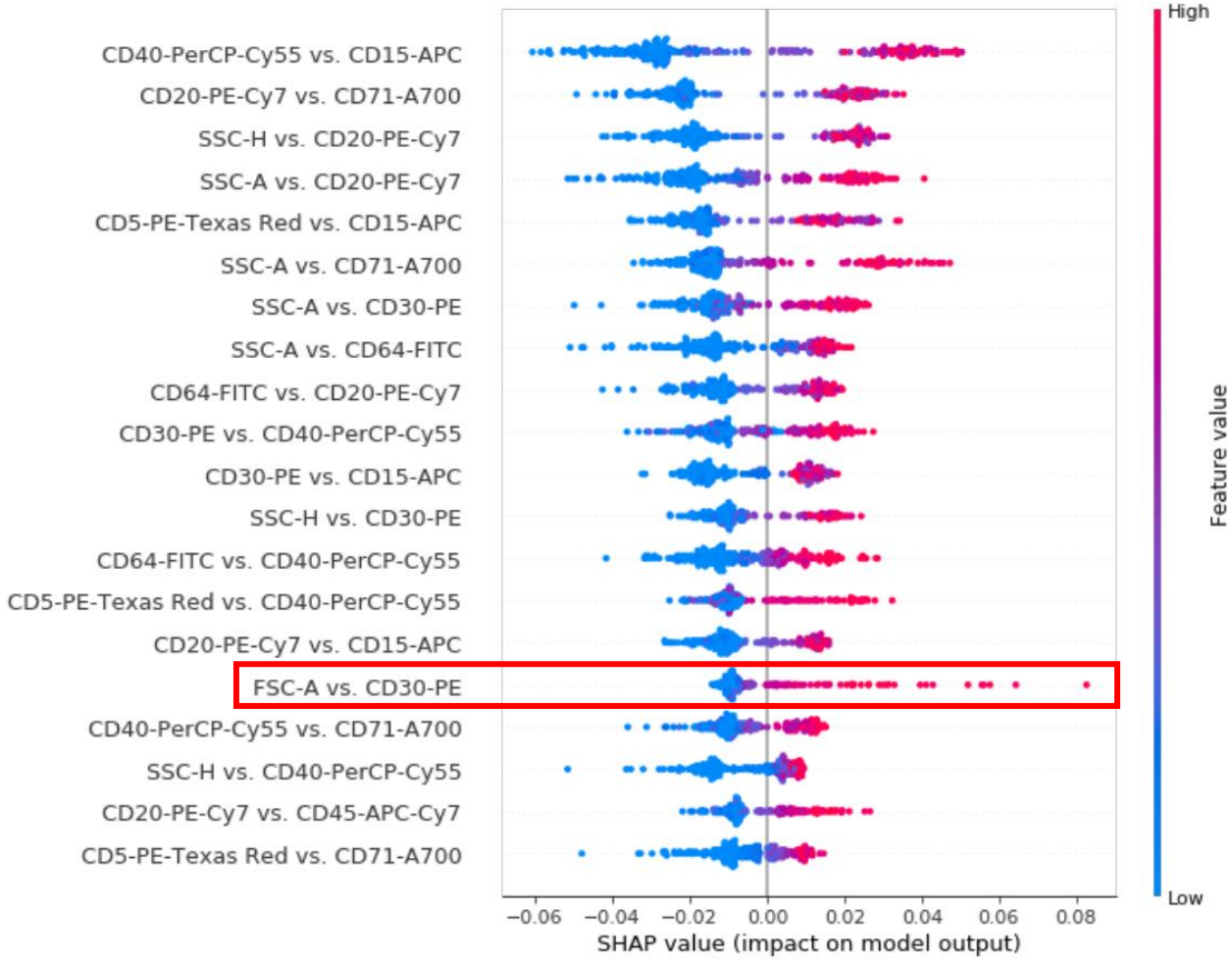
- SHAP: SHapley Additive exPlanations
- **Calculates the marginal impacts of each of the individual data elements on the final prediction scores**
- Can calculate values are for *individual* predictions
- Can be averaged to identify general trends



Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* **2**, 749–760 (2018).

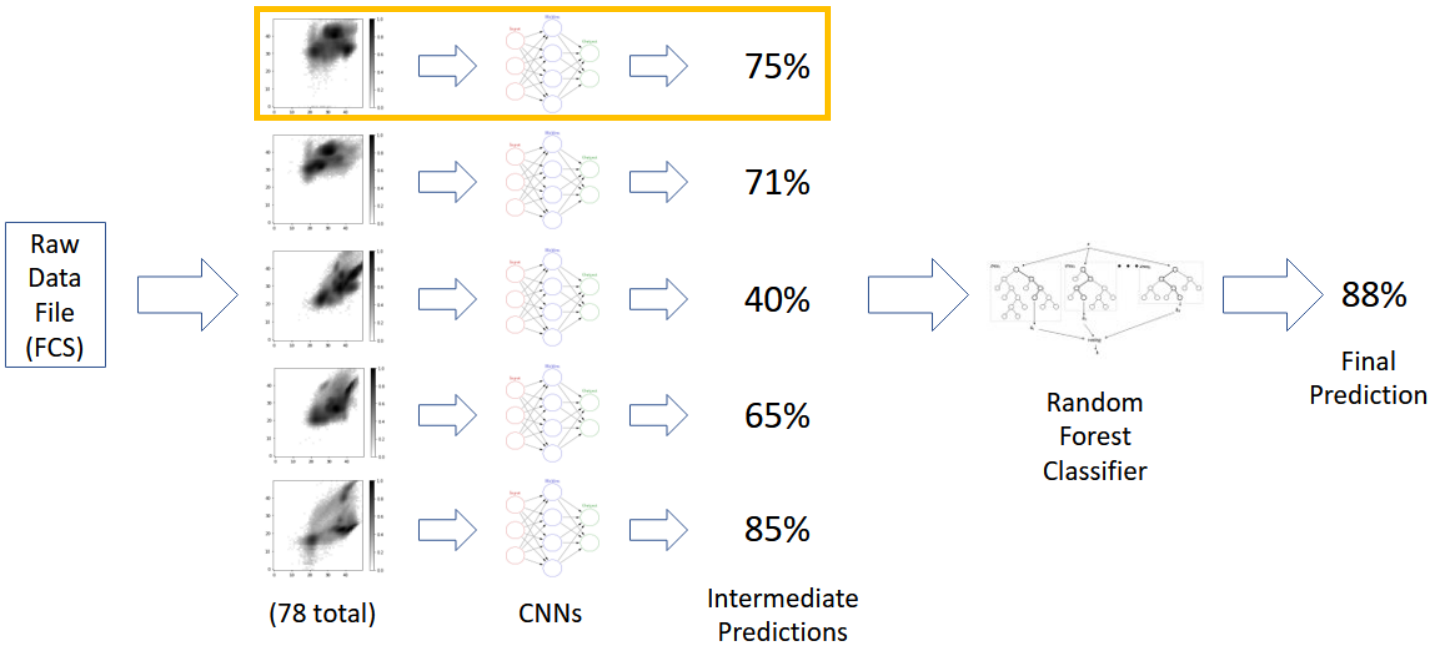
SHAP values highlight the most useful 2D histograms

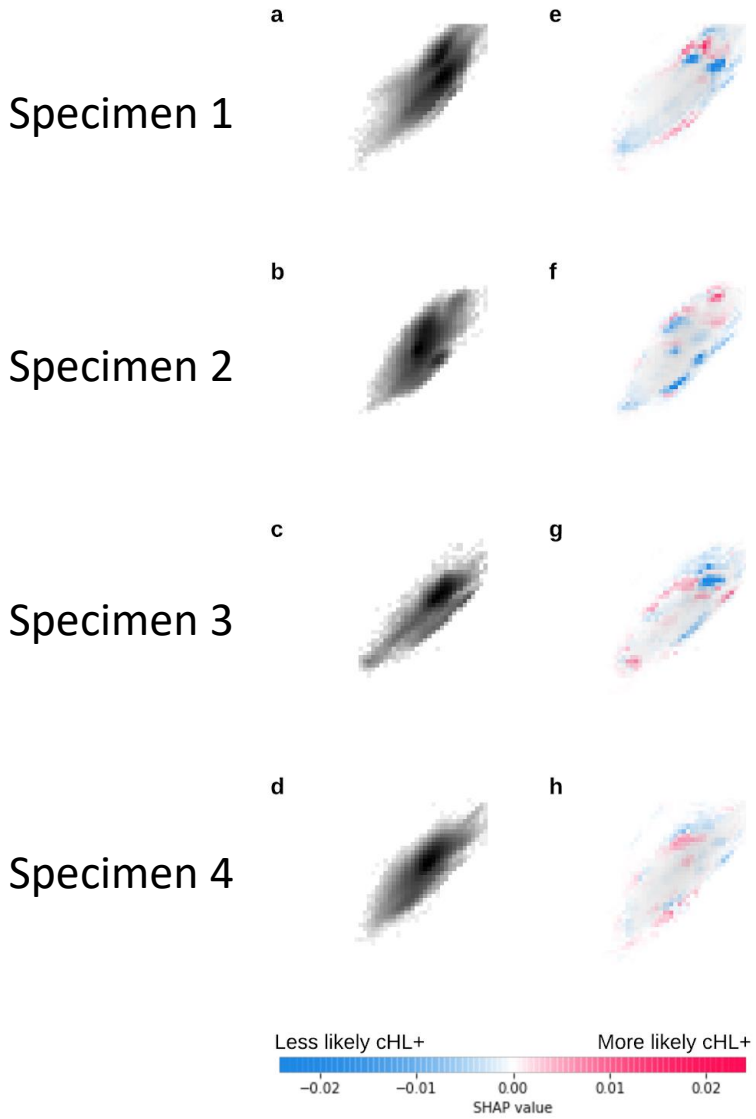




Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

SHAP values can also highlight key regions in individual histograms





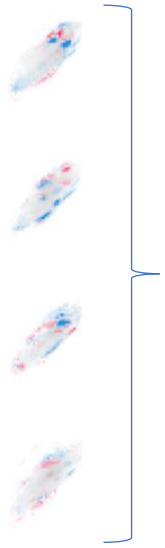
Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

That's great, but can you correlate that back to individual cells and plot it in normal flow cytometry software?

- Yes, we can!

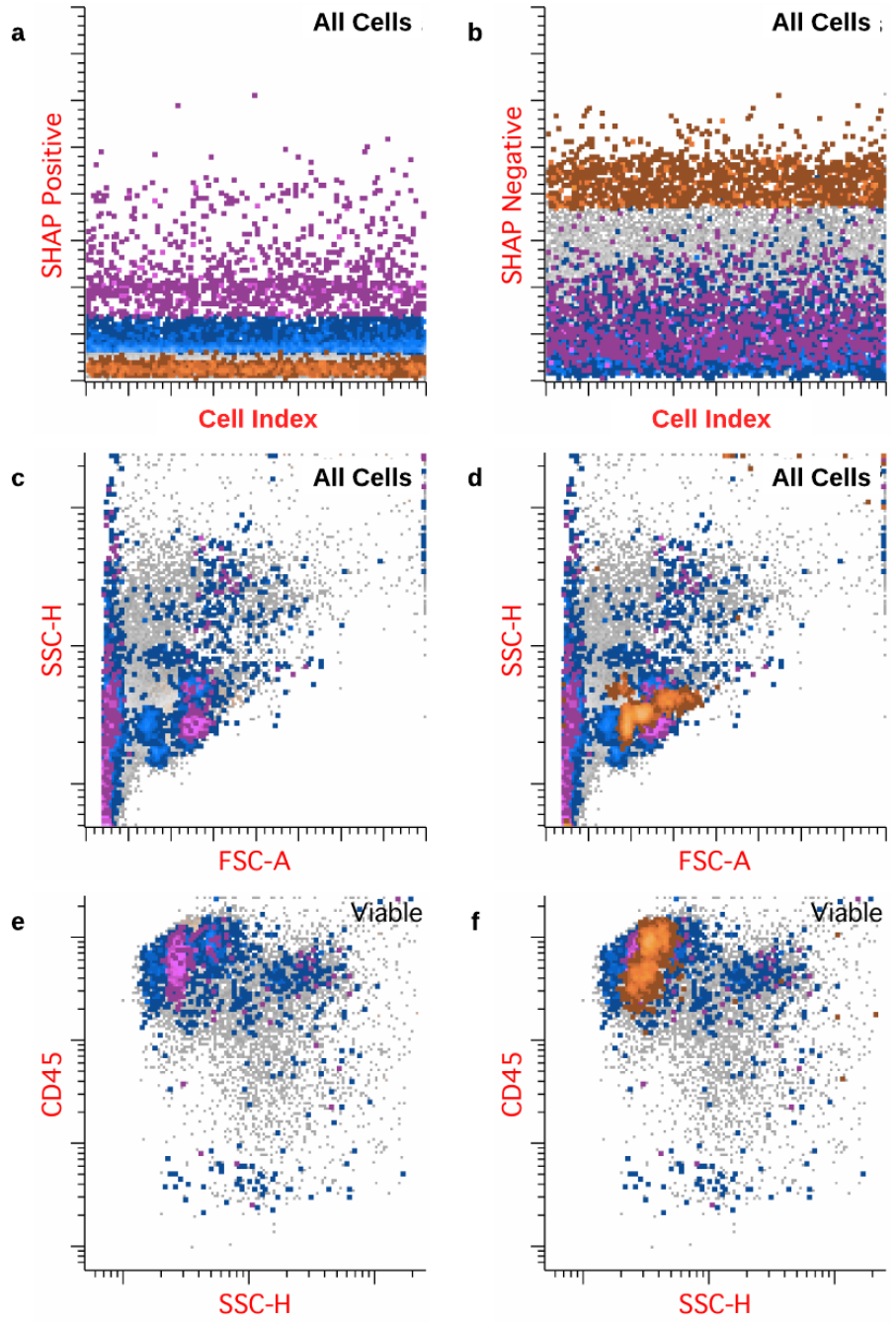
Visualizing important cell populations using standard software

- Integrate the SHAP values from each histogram bin within which an individual cell is found
- Save a new FCS file that includes the summed SHAP values for each cell



Raw Data File (FCS)

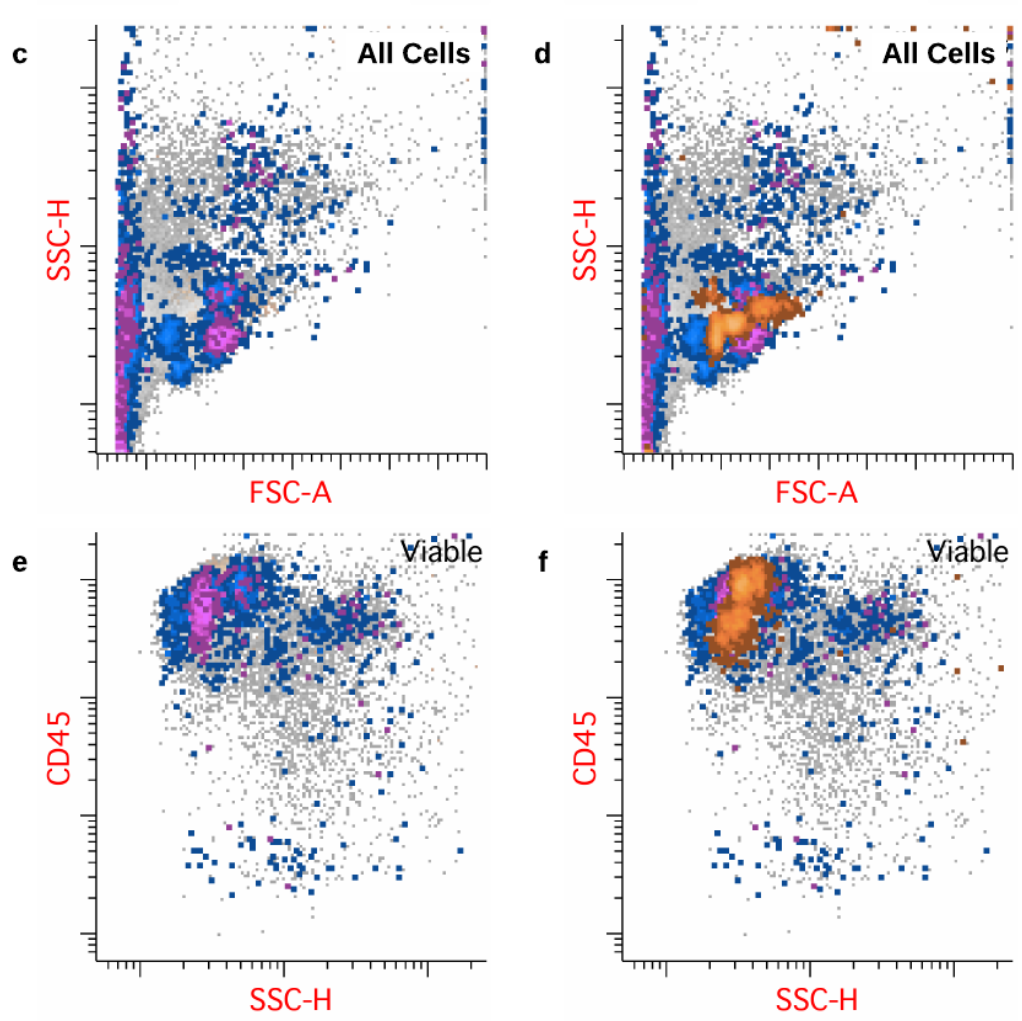
Blue = predictive of cHL+
Magenta = very predictive of cHL+
Orange = predicts against cHL+



Simonson PD, Wu Y, Wu D, Fromm JR, Lee AY. *Am J Clin Pathol.* 2021 Nov 8;156(6):1092-1102

What information it provides

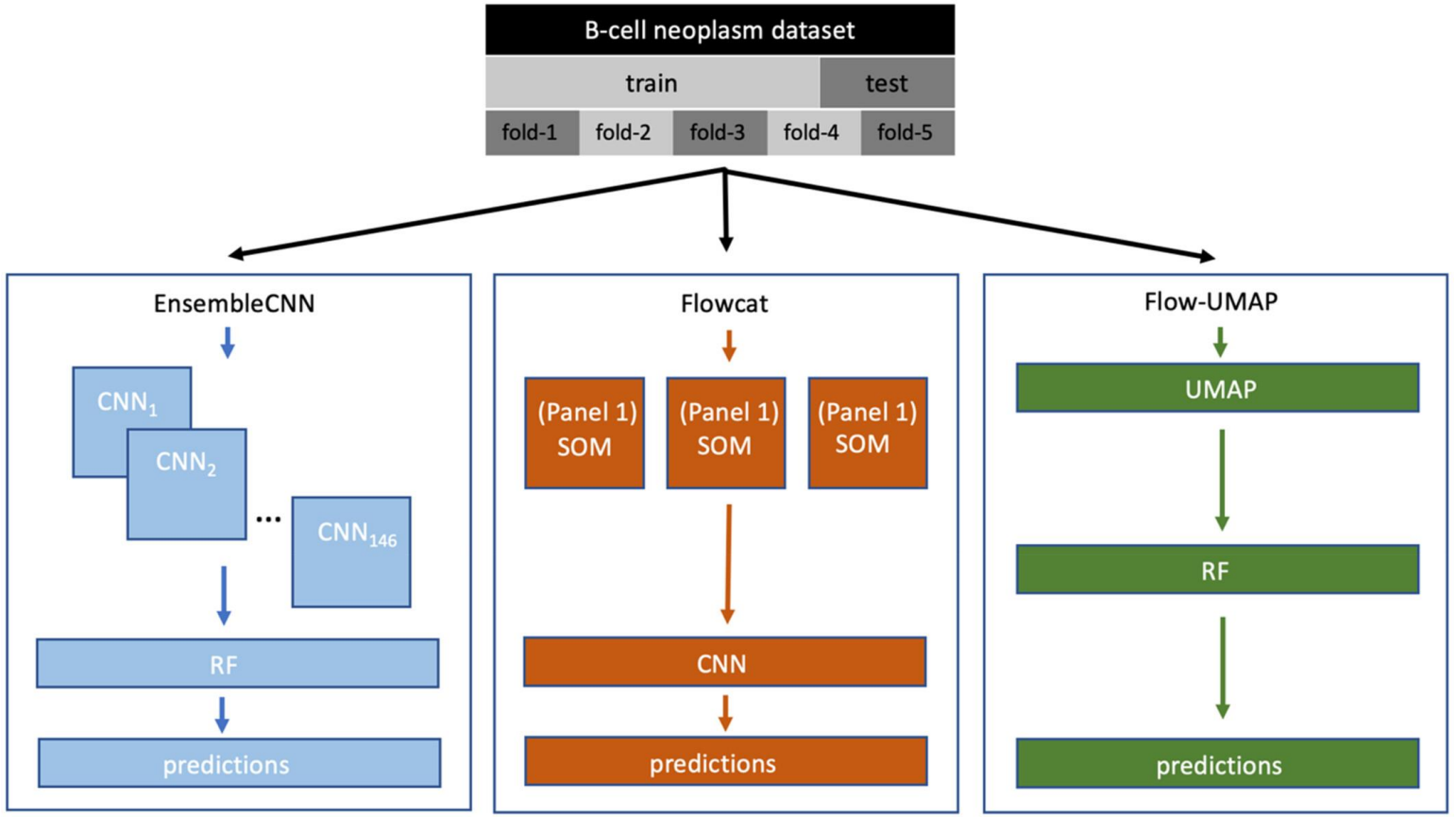
- Highlight cells that most impact the machine's decision making for a given case.
- Can potentially contribute to basic science by highlighting previously unrecognized associations.



Blue = predictive of cHL+
Magenta = very predictive of cHL+
Orange = predicts against cHL+

Which machine learning algorithm is best?

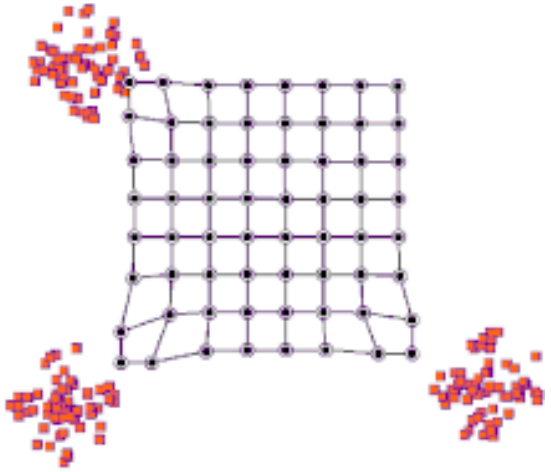
- Approach
  - Use public B cell neoplasms data set for comparison
- Classifiers compared
  - EnsembleCNN
  - FlowCat
  - UMAP-RF
- Comparisons
  - Accuracy
  - Speed
  - Interpretability



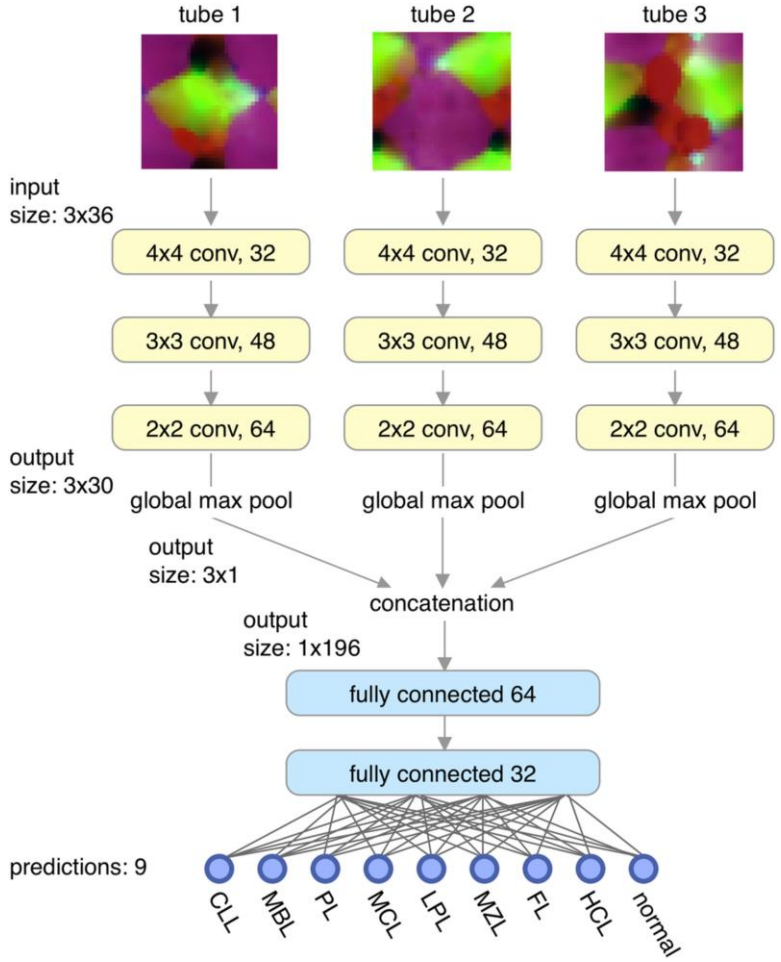
Dinalankara W et al. Cytometry B Clin Cytom. 2024 Jul;106(4):282-293.

FlowCat uses FlowSOM mappings as inputs for machine learning

- Generate FlowSOM mappings, which are passed to supervised ML algorithms
- Examples:
  - Identifying B cell neoplasms by machine learning (Zhao M et al. Cytometry A. 2020 Oct;97(10):1073-1080) (FlowCAT)
  - Identifying MDS (Duetz C et al. Cytometry A. 2021 Aug;99(8):814-824.)



https://en.wikipedia.org/wiki/Self-organizing_map



Zhao M et al. Cytometry A. 2020 Oct;97(10):1073-1080

B-cell neoplasms dataset

- Healthy cases and 9 sub-types of mature B-cell neoplasms
  - *Source: "Hematologist-Level Classification of Mature B-Cell Neoplasm Using Deep Learning on Multiparameter Flow Cytometry Data", Zhao et al., Cytometry, 2020.*
- 3 Tubes with 9-11 markers in each; 21152 total cases, 20731 cases after filtering

type	count	description
normal	11373	normal
CLL	4442	chronic lymphocytic leukemia
MBL	1614	monoclonal B-cell lymphocytosis
MZL	1121	marginal zone lymphoma
LPL	731	lymphoplasmacytic lymphoma
PL	593	prolymphocytic leukemia
MCL	313	mantle cell lymphoma
FL	250	follicular lymphoma
HCL	225	hairy cell leukemia
HCLv	69	hairy cell leukemia variant

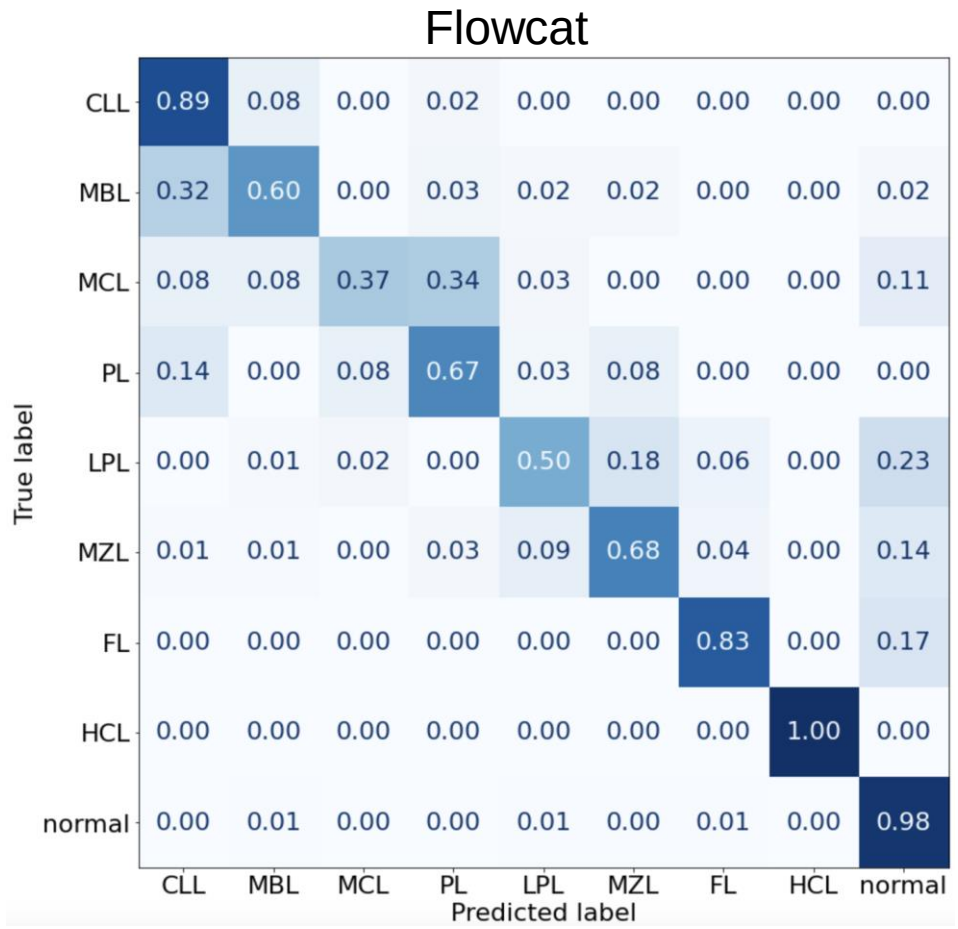
	Panel 1	Panel 2	Panel 3
1	'FS INT LIN'	'FS INT LIN'	'FS INT LIN'
2	'SS INT LIN'	'SS INT LIN'	'SS INT LIN'
3	'FMC7-FITC'	'Kappa-FITC'	'CD8-FITC'
4	'CD10-PE'	'Lambda-PE'	'CD4-PE'
5	'IgM-ECD'	'CD38-ECD'	'CD3-ECD'
6	'CD79b-PC5.5'	'CD25-PC5.5'	'CD56-APC'
7	'CD20-PC7'	'CD11c-PC7'	'CD19-APCA750'
8	'CD23-APC'	'CD103-APC'	'HLA-DR-PacBlue'
9	'CD19-APCA750'	'CD19-APCA750'	'CD45-KrOr'
10	'CD5-PacBlue'	'CD22-PacBlue'	
11	'CD45-KrOr'	'CD45-KrOr'	

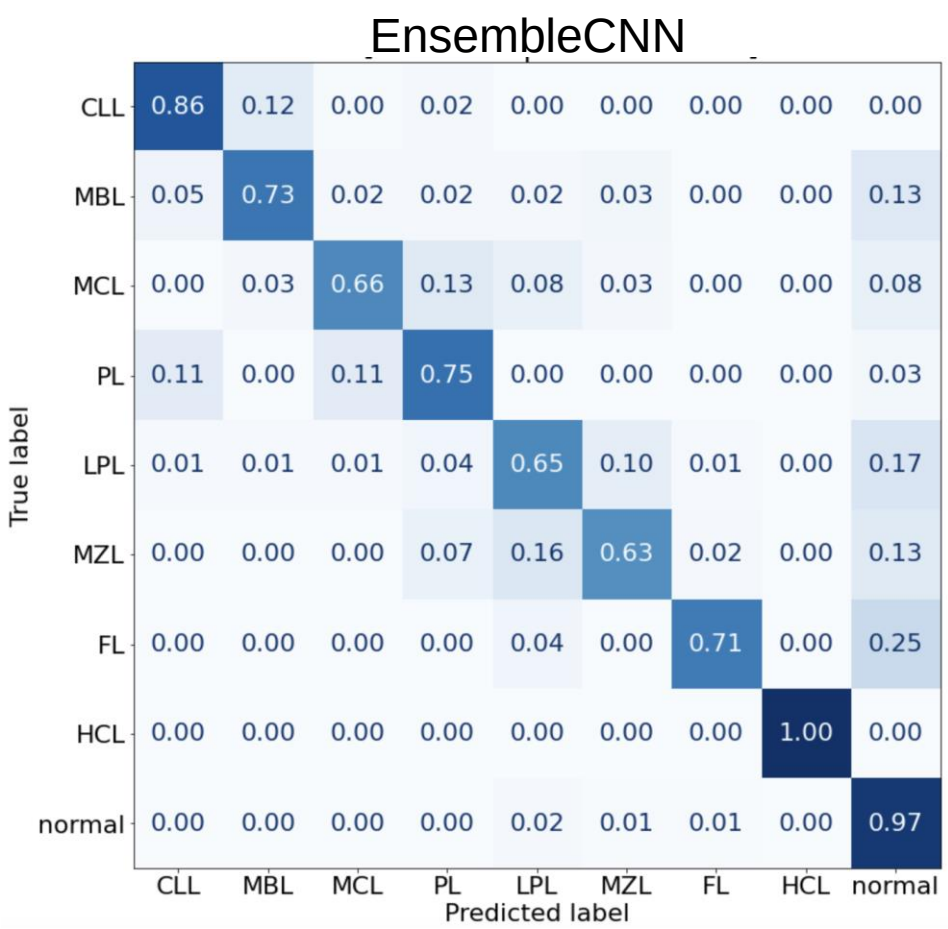
Training/Testing Data Sets

group	train (CNN)	train (RF)	test
CLL	2016	2019	403
MBL	776	778	60
MCL	136	139	38
PL	278	277	36
LPL	320	322	84
MZL	498	499	116
FL	110	112	24
HCL	100	101	24
normal	4900	4903	1563

- We followed a similar training/testing split as used in Zhao et. al. (FlowCat)
- For EnsembleCNN, the training data was split further into 50% for CNN training and 50% for the random forest

Accuracy is very similar between the two





	Precision		Recall		F1-score	
	Flowcat	EnsembleCNN	Flowcat	EnsembleCNN	Flowcat	EnsembleCNN
macro average	0.70	**0.72**	0.72	**0.77**	0.70	**0.74**
weighted average	0.91	**0.92**	**0.90**	**0.90**	**0.91**	**0.91**

Dinalankara W et al. Cytometry B Clin Cytom. 2024 Jul;106(4):282-293.

Time required to train the classifiers

Flowcat (python 3.6 with GPU)

- ~ 15 hrs, 45 mins
- Almost all time is used for SOM creation

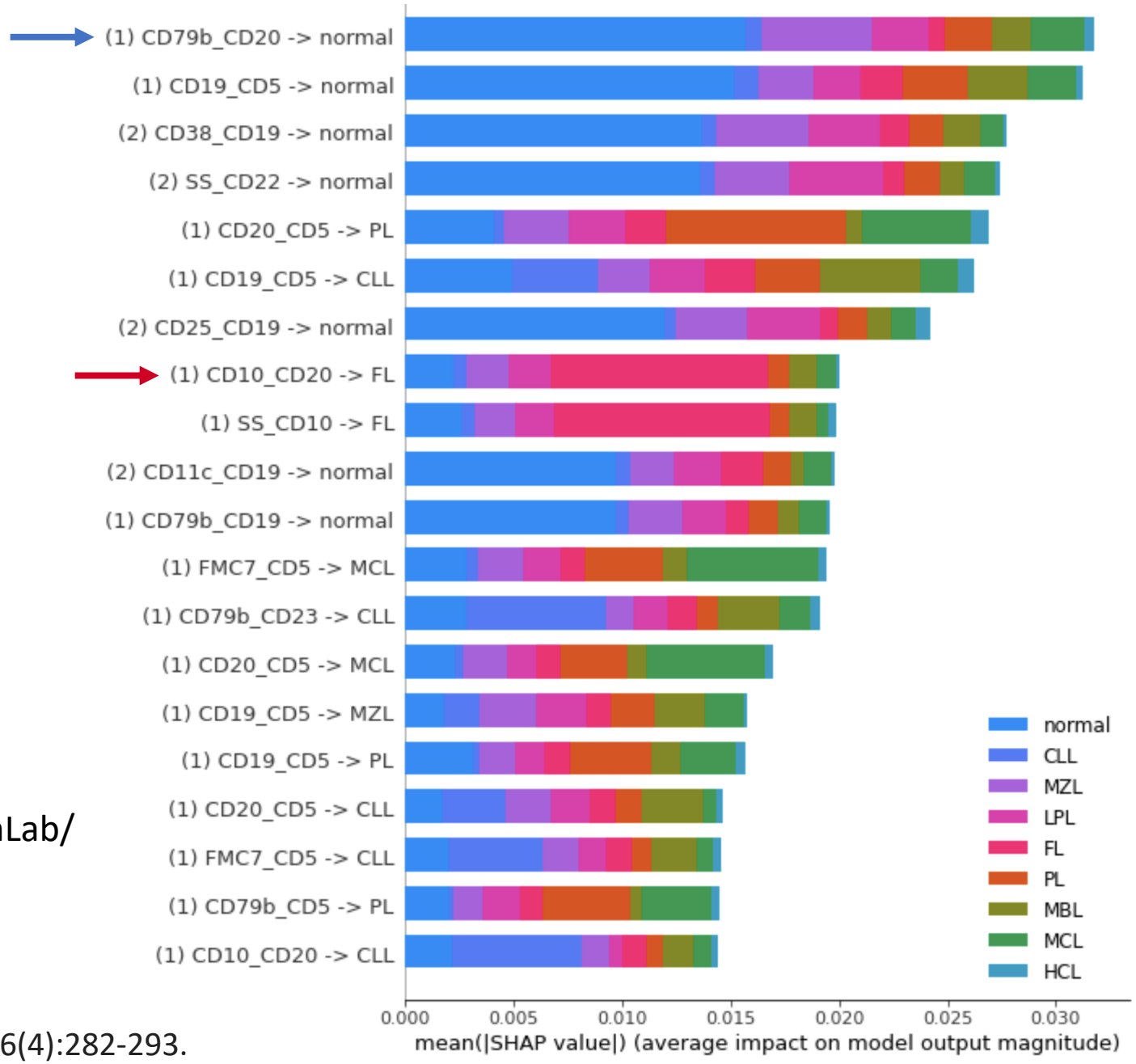
EnsembleCNN (python 3.9 with GPU)

- ~ 5 hr, 30 mins
- ~4 hours for creating case histograms for all the cases (mostly I/O operations)
- ~ 1 hr, 30 mins for training and testing the CNN+RF
- About one CNN per minute

Dinalankara W et al. Cytometry B Clin Cytom. 2024 Jul;106(4):282-293.

Interpretability

- SHAP (Shapley Additive Explanations) applied to EnsembleCNN



Python source code:
https://github.com/SimonsonLab/flowComparison

Dinalankara W et al. Cytometry B Clin Cytom. 2024 Jul;106(4):282-293.

Example application #2: Using machine learning to predict the need for additional antibody panels

- **Could reduce number of times a tech/pathologist looks at a case**
- CLL antibody panel helps distinguish CLL from mantle cell lymphoma, etc.

Data: UW B cells antibody panel (CD20, kappa, lambda, CD5, CD19, CD38, CD10, and CD45)

- Non-gated data (for now)
- First specimen only for each patient
- 10495 cases (~3 years' data)
  - 9.1% were positive for follow up CLL panel
- 80:20 training:test data

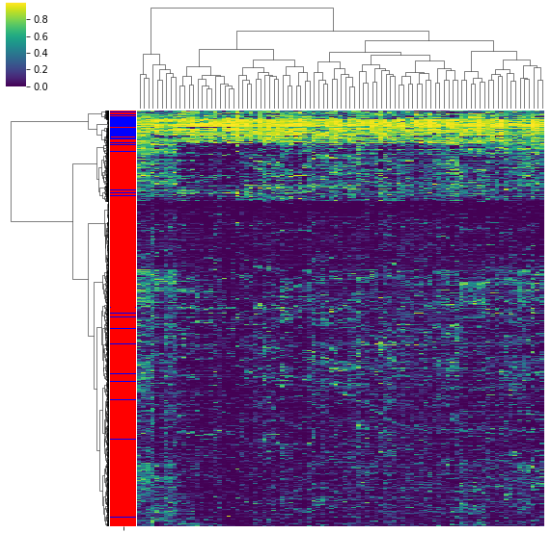
Results

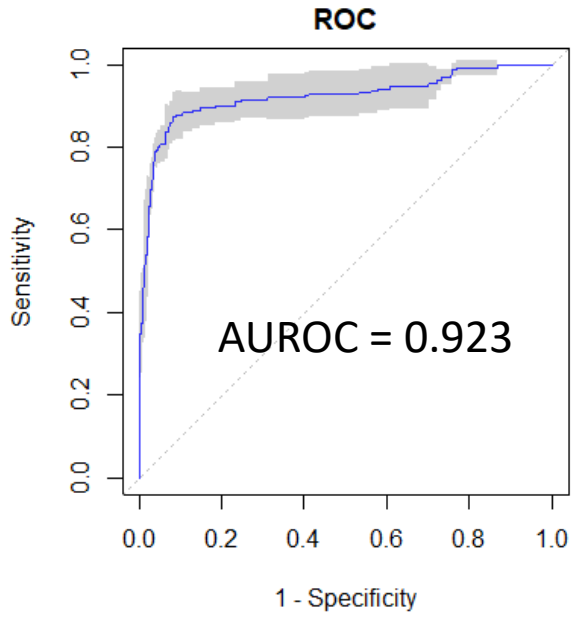
CLL panel ordered	True +	True -
Predicted +	140	51
Predicted -	50	1858

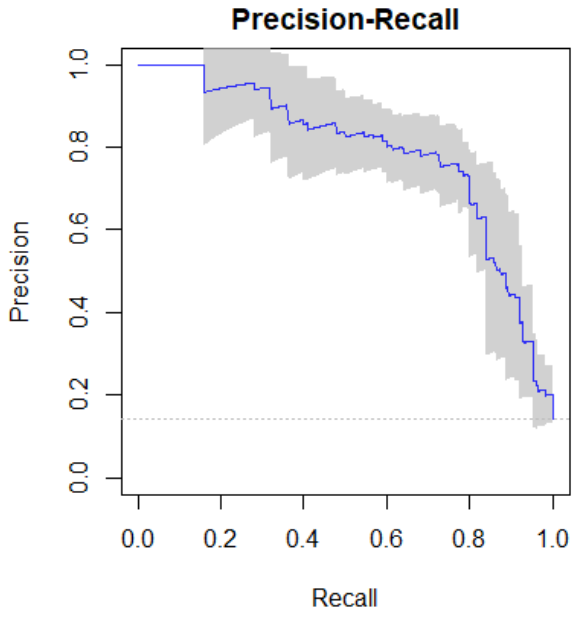
Accuracy = 95%

Limitations:

1. Data were only "roughly compensated"
2. Classifier is not aware of patient history (CLL add-on is not ordered if patient is already known to have CLL).





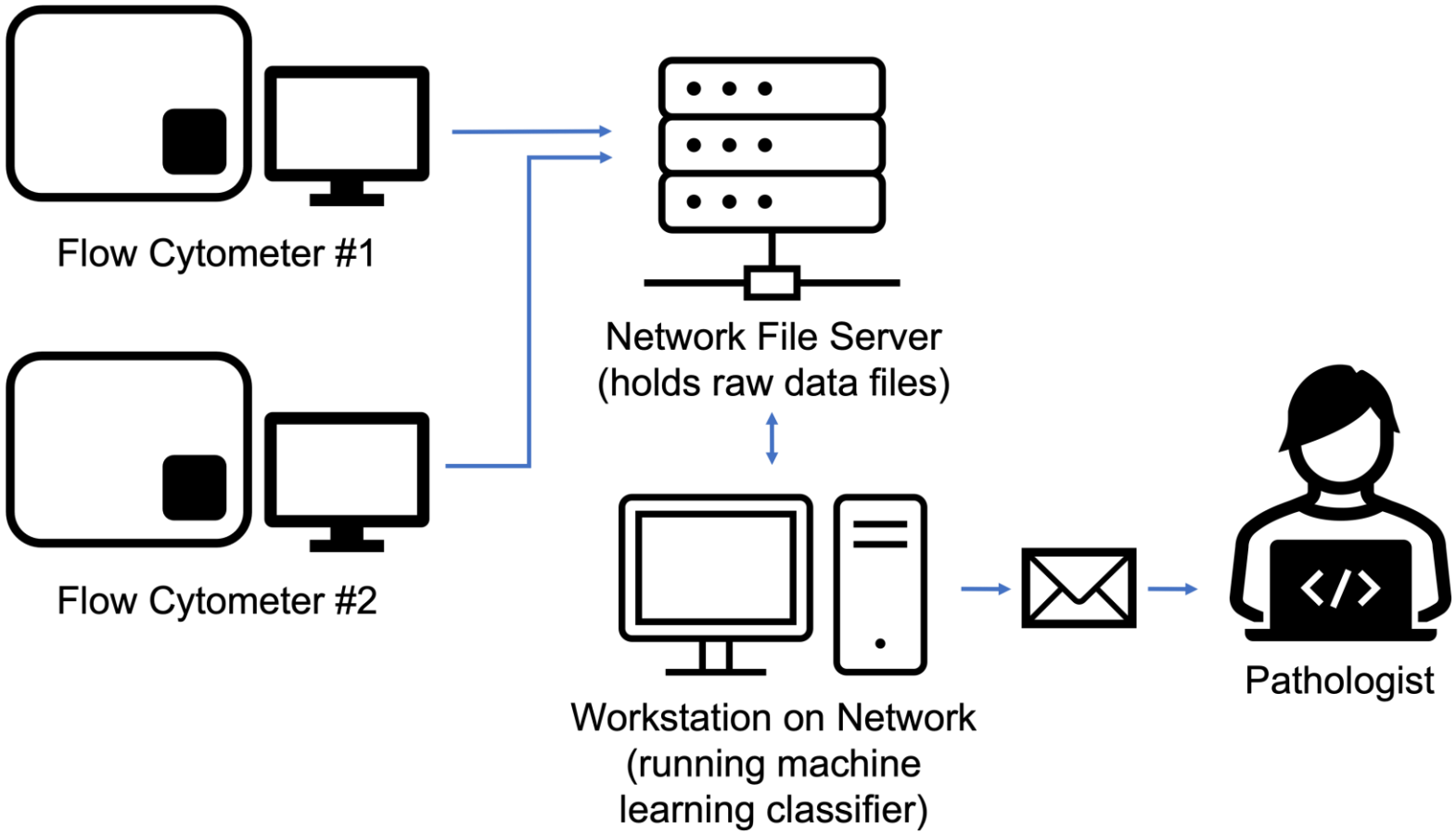


That's great, but when are you going to actually apply it to real, prospective data with real-time analysis?

- We already did!

System for real-time prediction and alerting for adding CLL antibody panel

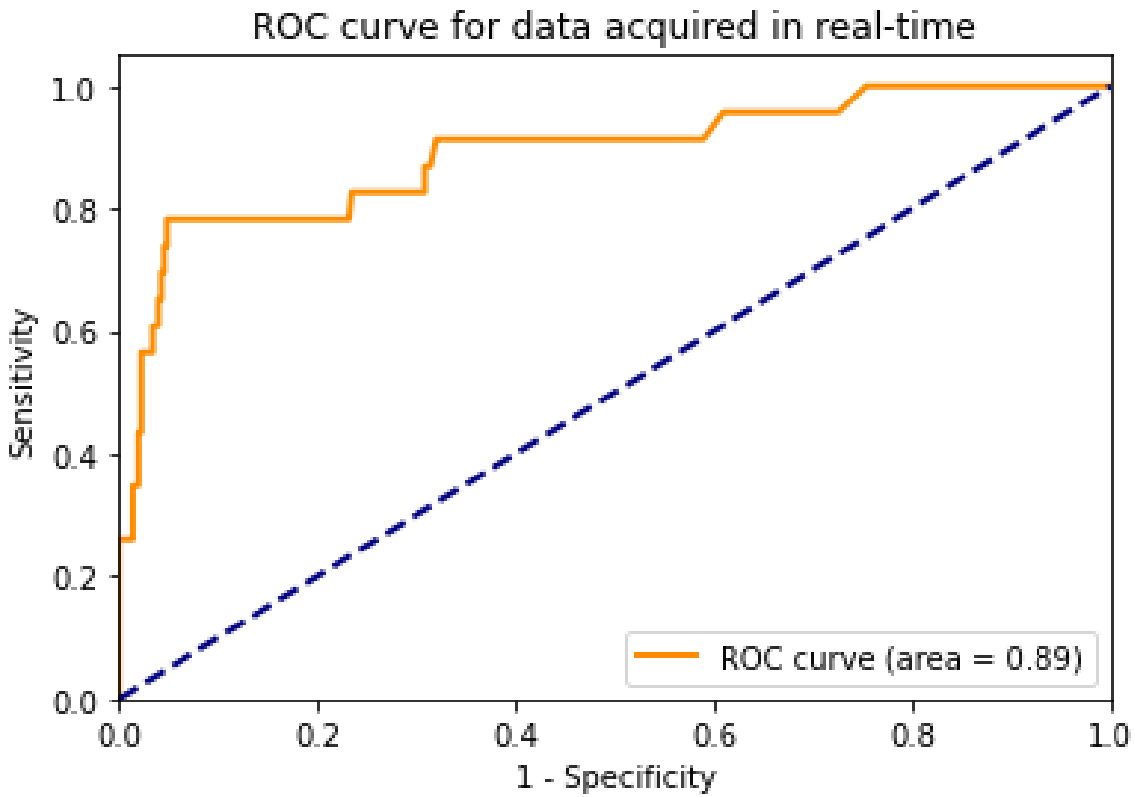
- Workstation checks the FCS file server every minute for newl B cell panel FCS file.
- Each FCS file receives a probability score.
- If probability meets determined threshold, the computer sends an alert email with accession number.

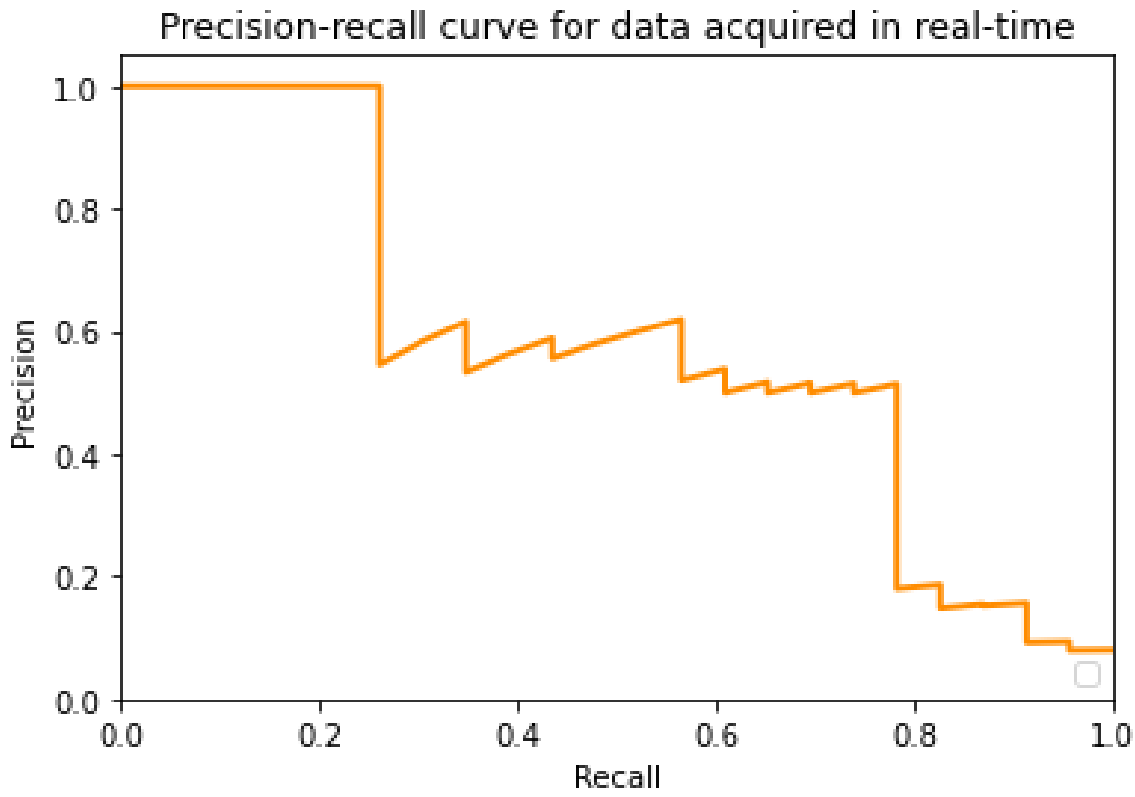


Simonson PD, Lee AY, Wu D. *Am J Clin Pathol.* 2021 Nov 1

Prospective, real-time application of the classifier

- 367 prospective cases during March-May, 2020, 6.1% positive for CLL add-on.





Simonson PD, Lee AY, Wu D. *Am J Clin Pathol.* 2021 Nov 1

Inspecting the cases the system got wrong

	Predicted no CLL order	Predicted CLL order
CLL panel not ordered	336	17 (false pos.)
CLL panel ordered	5 (false neg.)	18

On closer examination...

Of the 5 false negative predictions:

- 3 cases did not have CLL panel included in final report (2 normal and 1 suspicious for CD5-negative DLBCL).
- 1 case suspicious for a small number of CD5-negative large B cell lymphoma cells.
- 1 case was CSF involved by mantle cell lymphoma (CLL panel was justified).

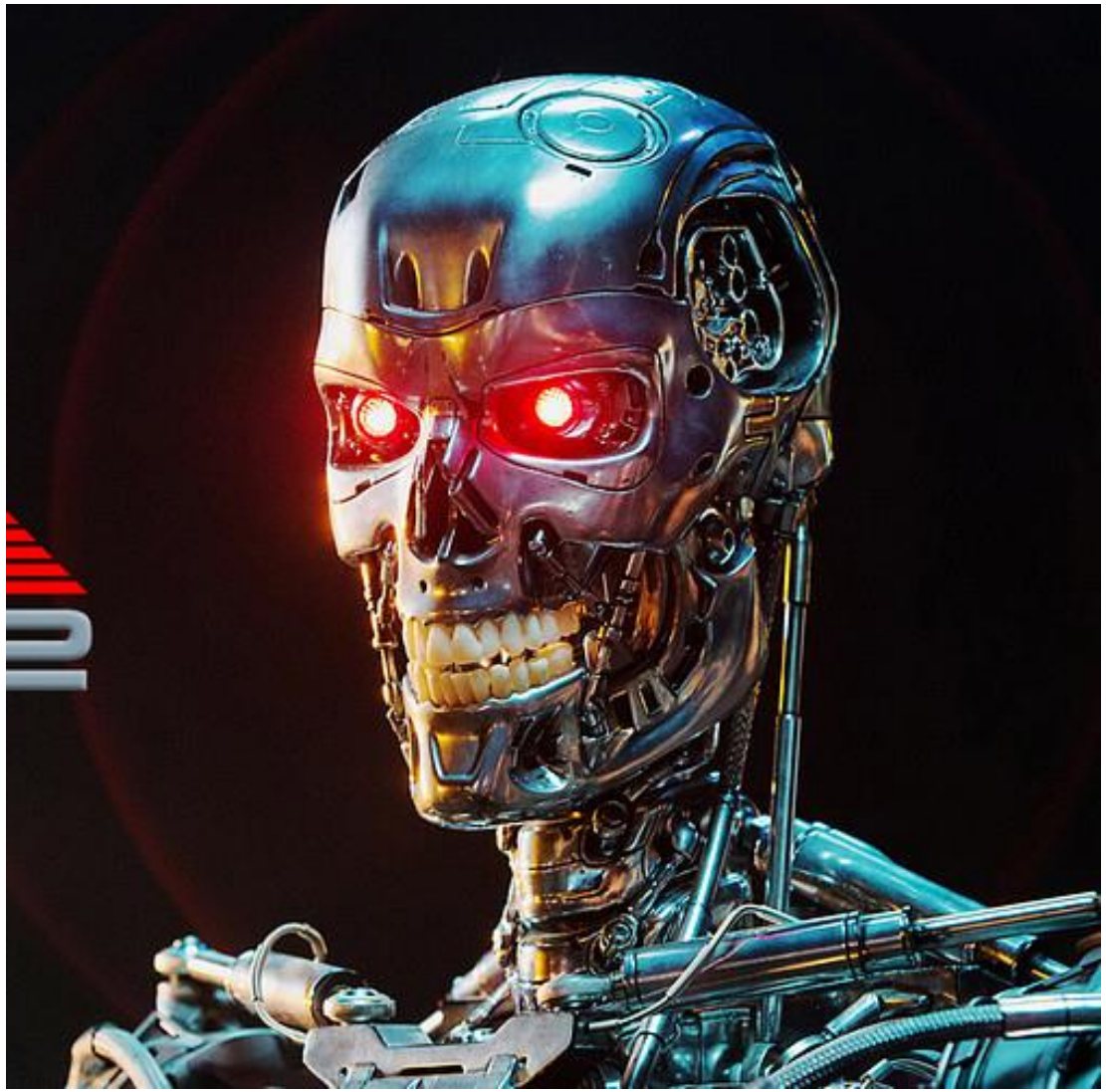
Of the 17 false positive predictions:

- 13 had a prior diagnosis of CLL/SLL and 1 had a prior diagnosis of mantle cell lymphoma, so CLL panel was not ordered per lab protocol.
- 3 were CD5-negative B cell lymphomas.

Simonson PD, Lee AY, Wu D. *Am J Clin Pathol.* 2021 Nov 1

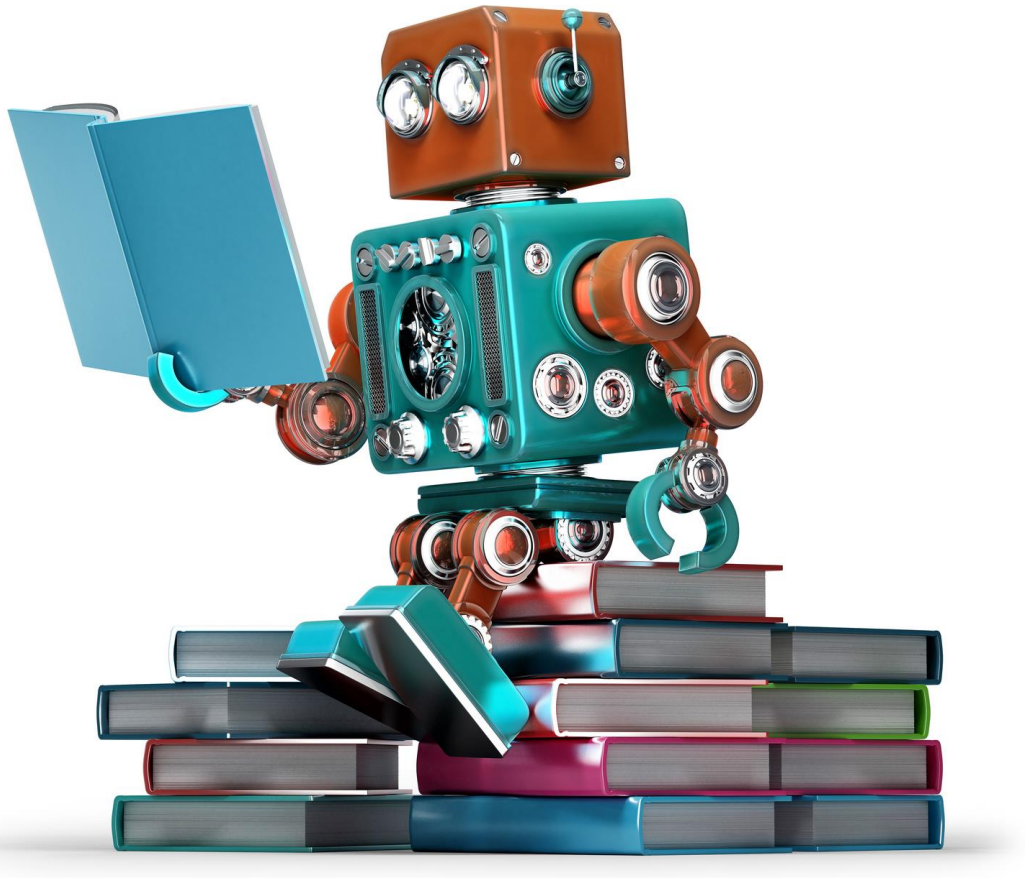
Challenges in applying machine learning to flow cytometry

- Regulatory uncertainty
- Data availability
- Generalizability of algorithms



Summary

- There are many potential applications for machine learning/AI in flow cytometry.
- Machine learning can assist in identifying diagnoses, predicting add-on tubes, etc.
- Interpretability tools can help in understanding how the algorithms are functioning with respect to data.
- Don’t overlook easy ways to improve efficiency (e.g., rules-based report generation).



Acknowledgments

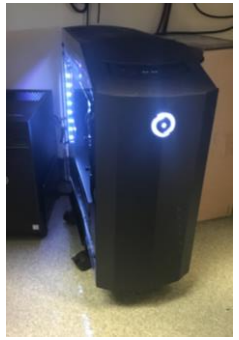
- Luigi Marchionni
- Sophie Rand
- Wikum Dinalankara
- Sanjay Patel
- David Brundage
- Nadine Demko
- Tayler Van den Akker
- James Yip

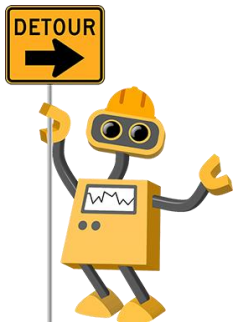
- Yue Wu
- David Wu
- Jonathan Fromm
- Aaron Lee

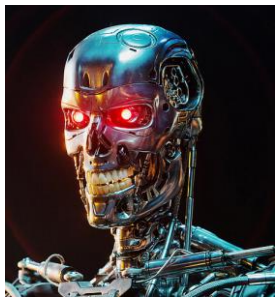
- David Ng

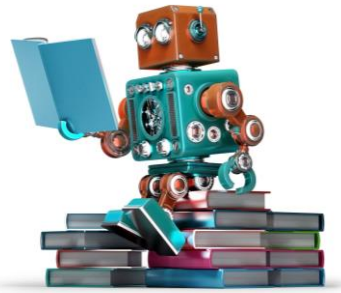
Funding

- Weill Cornell Medicine Department of Pathology and Laboratory Medicine
- University of Washington Department of Laboratory Medicine
- Roger Moe Fellowship (PDS)
- Donation of a GPU by NVIDIA Corporation (PDS)
- NEI K23EY029246 (AYL)
- Research to Prevent Blindness (AYL)









Questions?