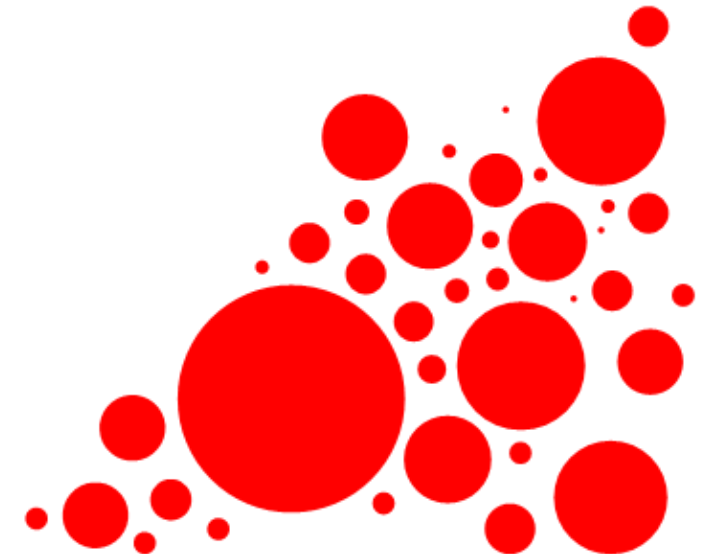




# **Application of artificial intelligence models to flow cytometric data for diagnosing hematologic neoplasms**

Wolfgang Kern

MLL Munich Leukemia Laboratory, Munich, Germany

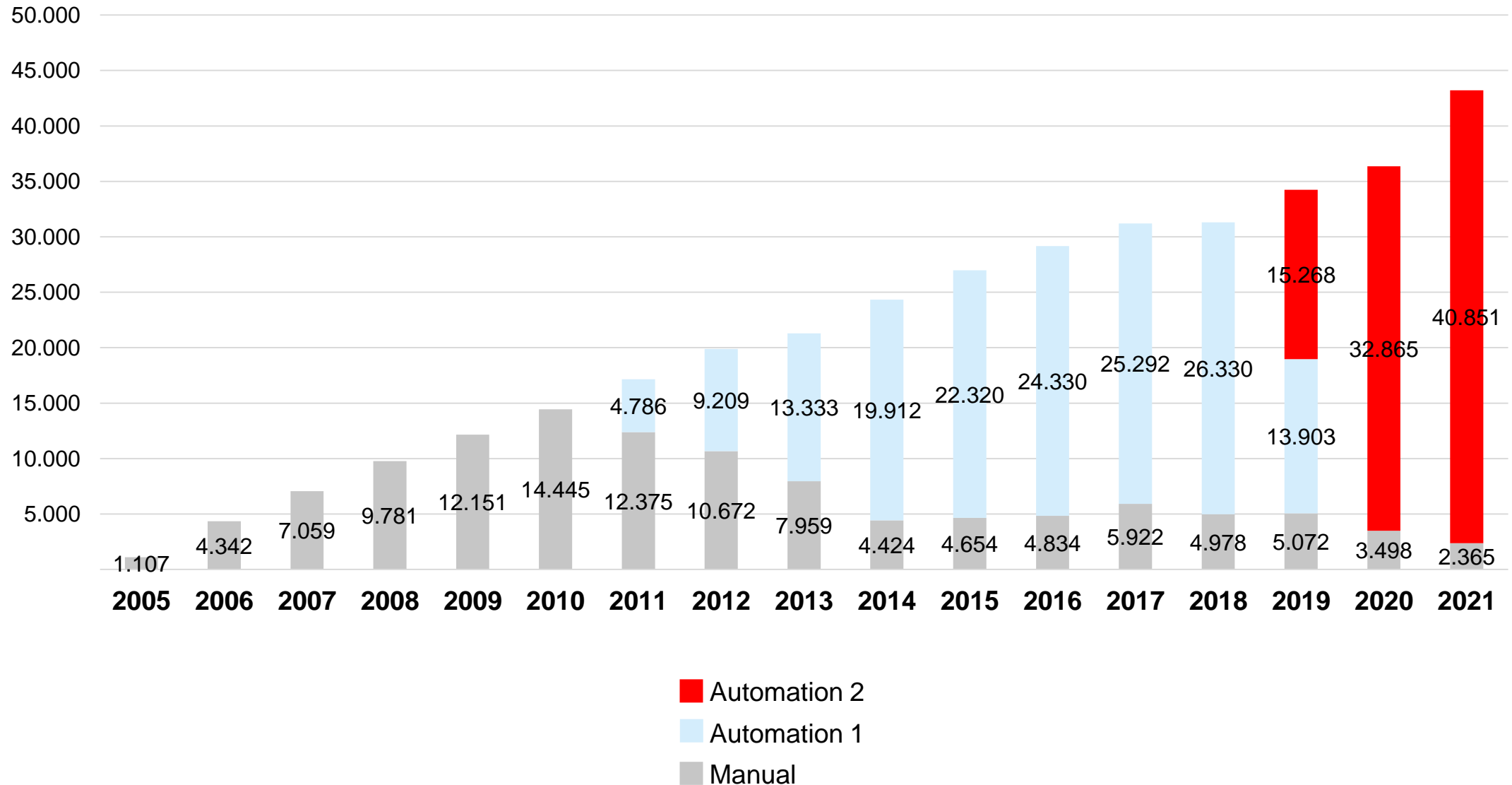


# Conflict of Interest Disclosure

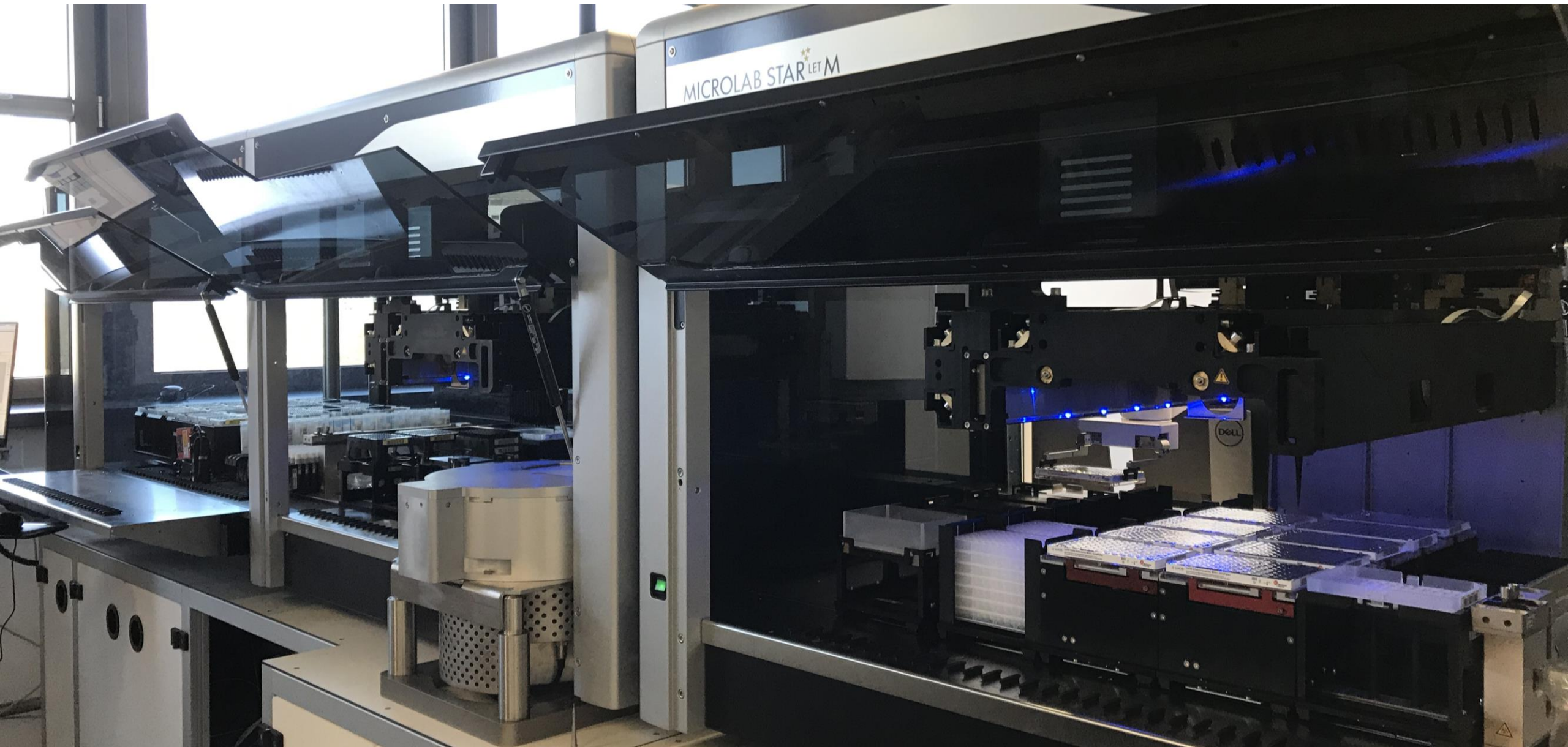


WK is part owner of MLL Munich Leukemia Laboratory

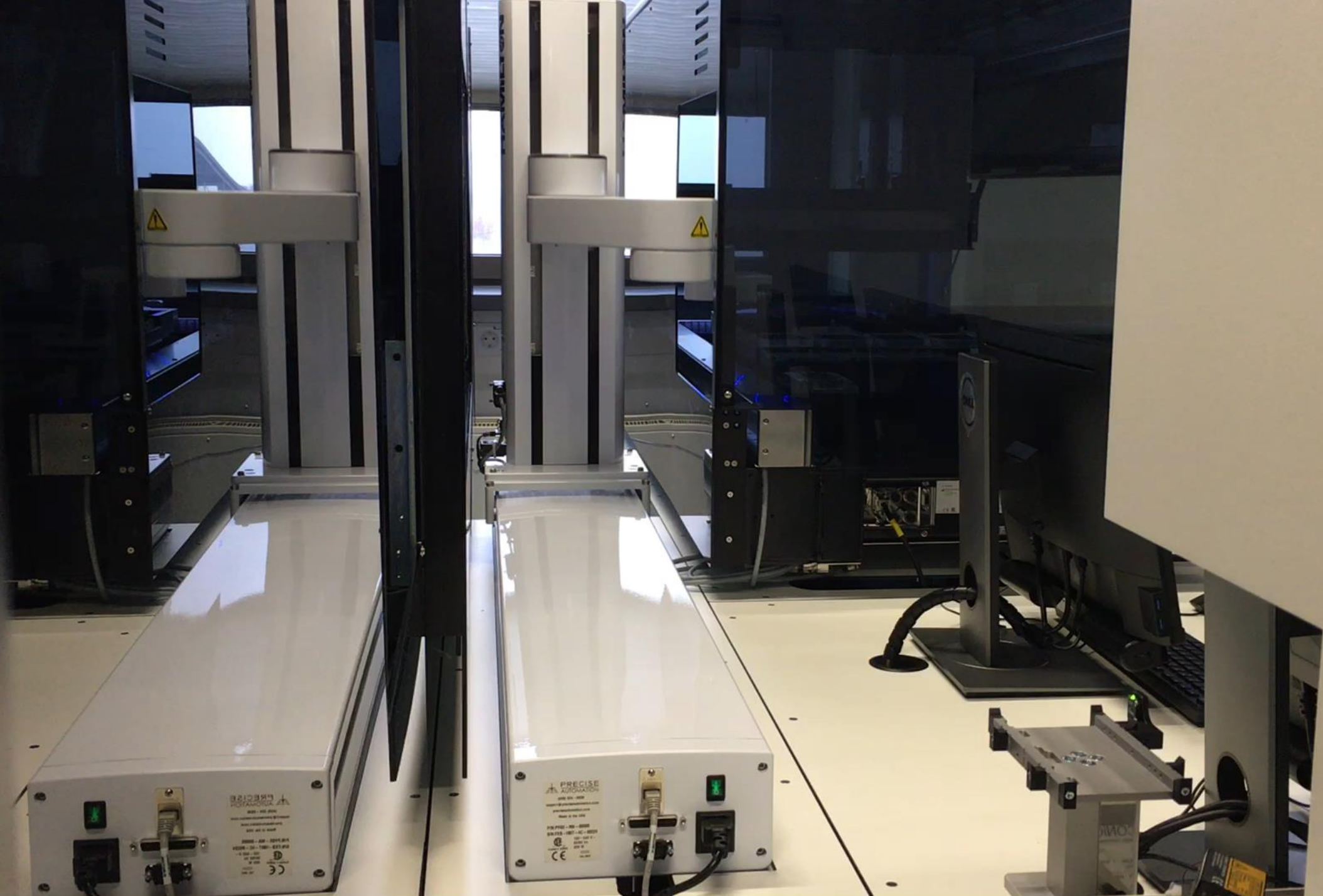
# Workload in the flow lab



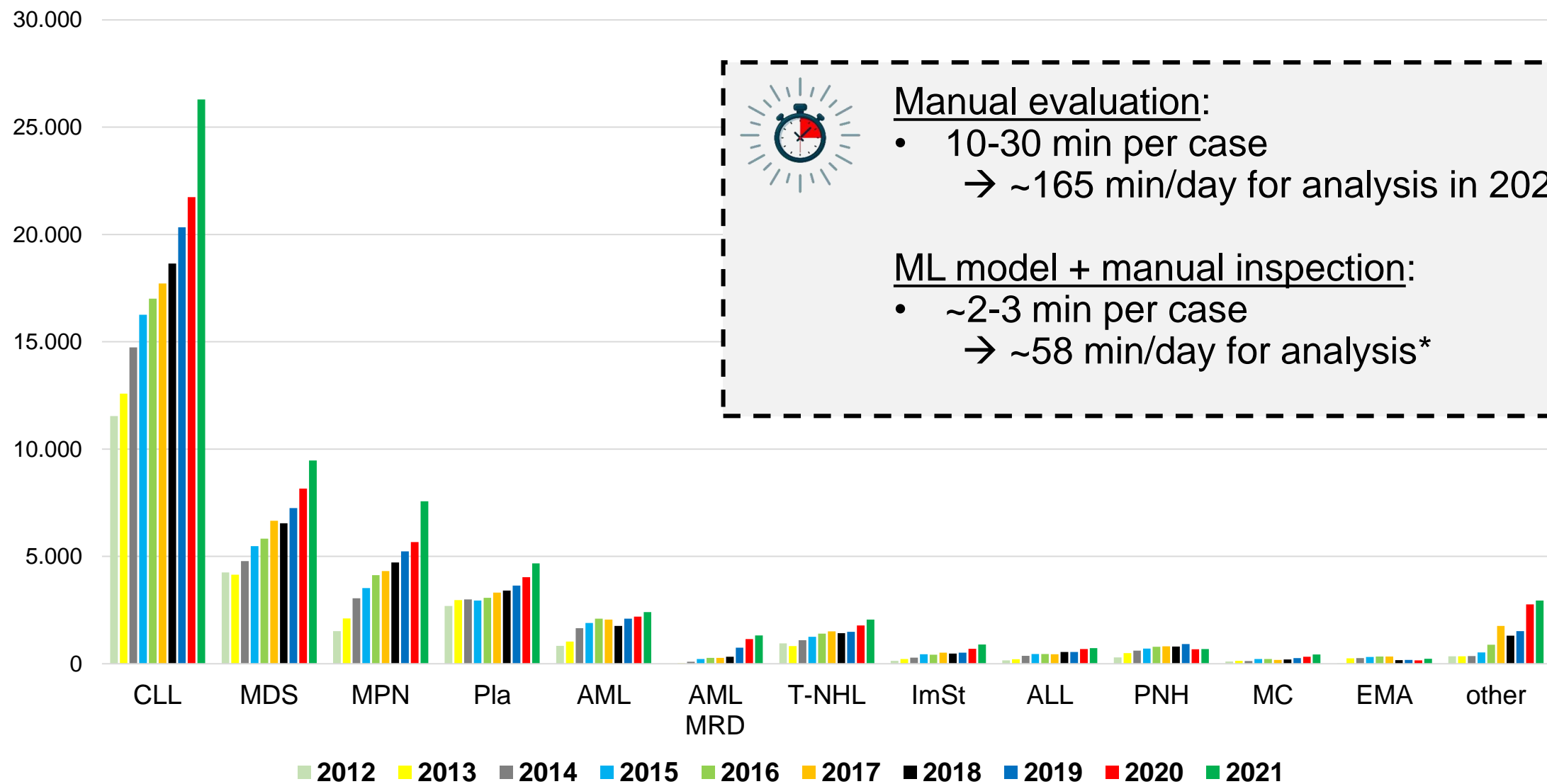
# Workload in the flow lab

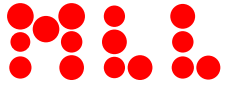




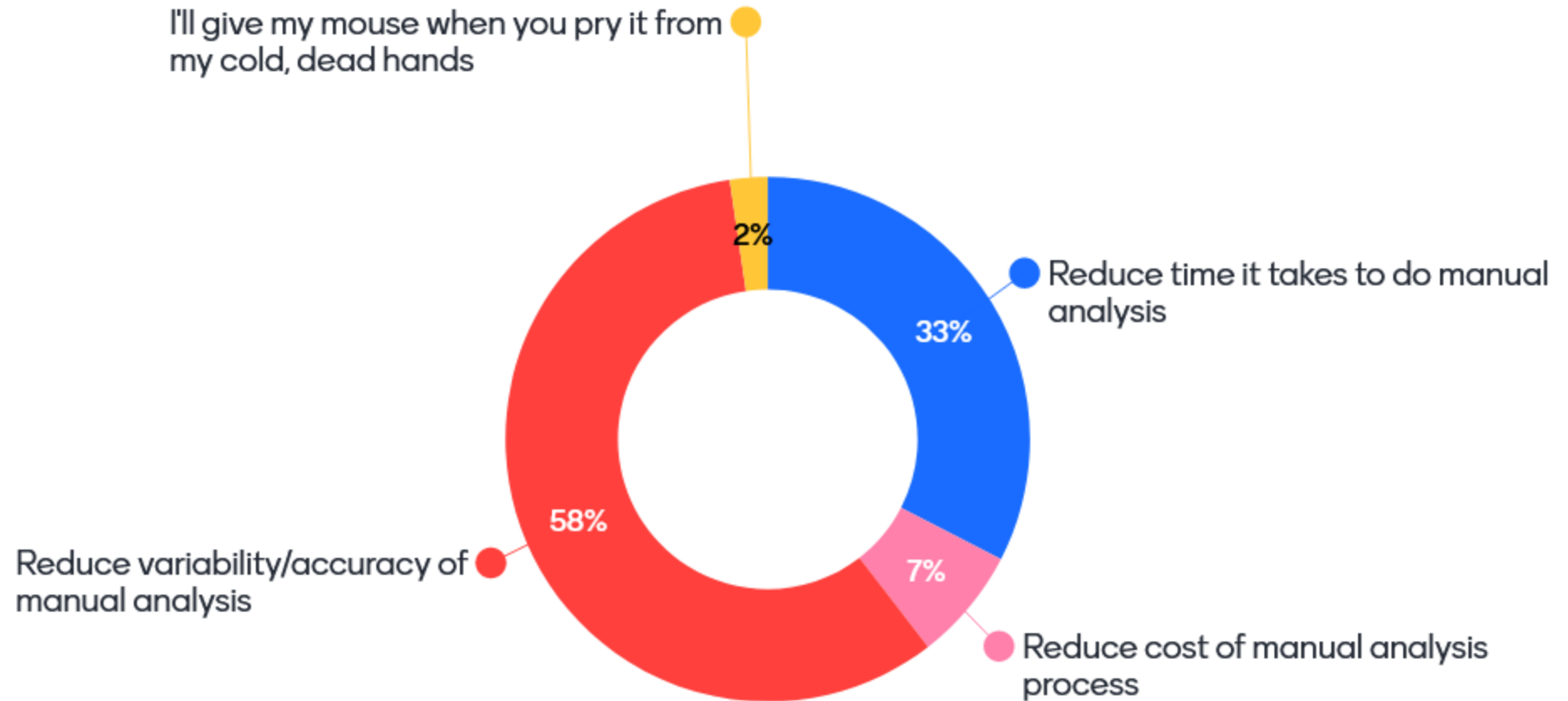


# Numbers of panels per year



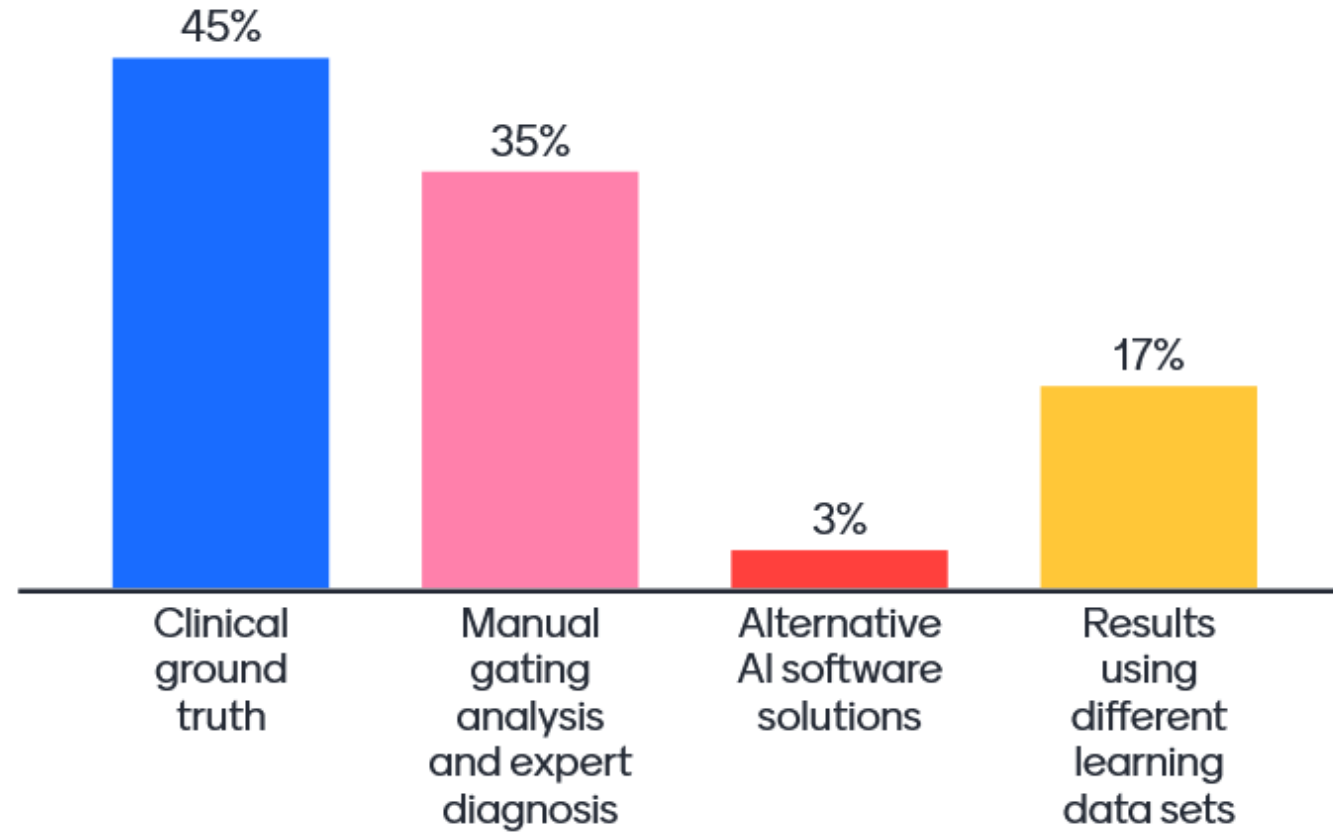


# Which of these is the most looked for benefit of ML approaches to aid clinical data analysis?



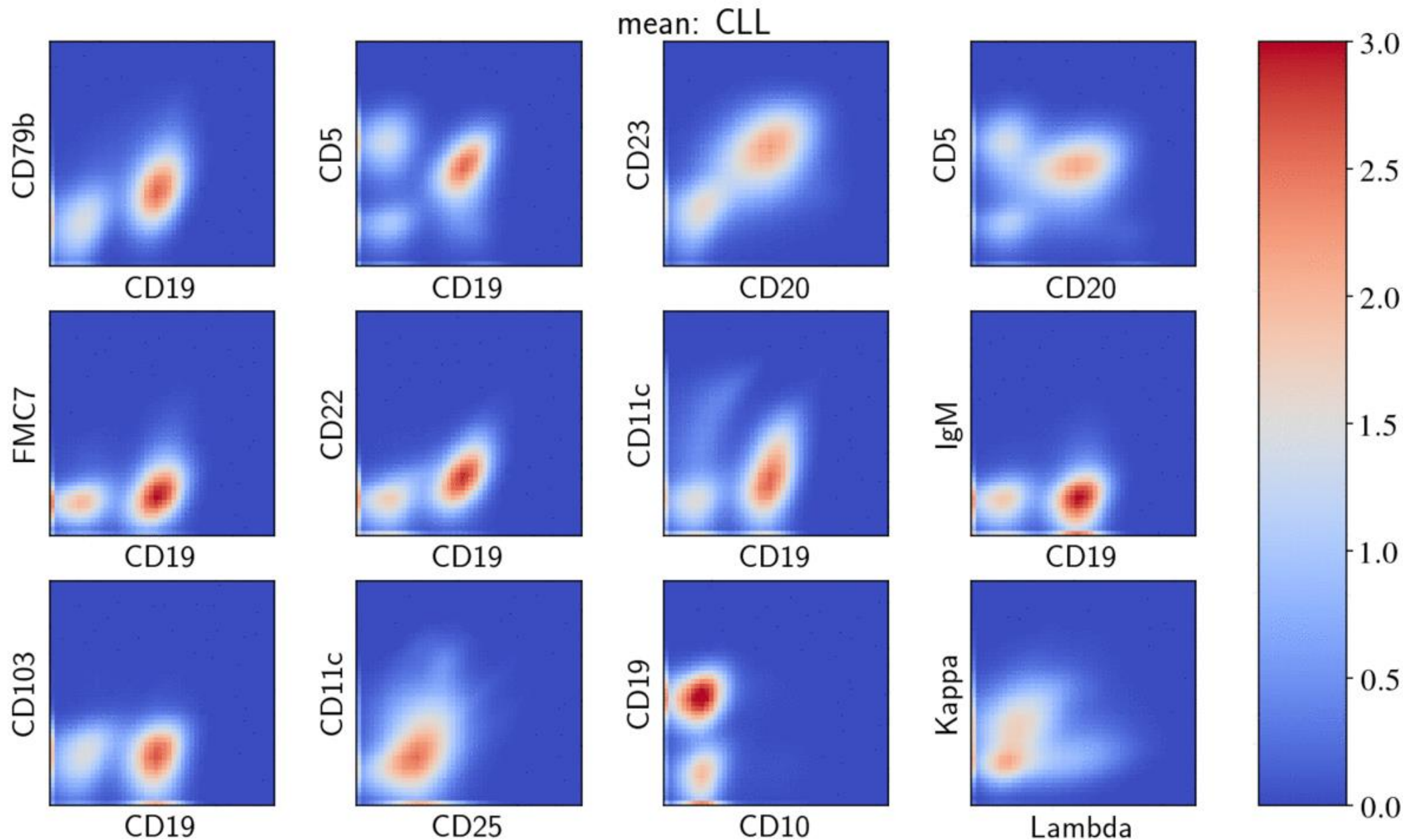
# Which tests would you find essential to perform for reliable validation?

## Comparison with:

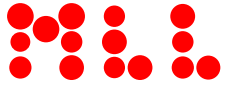




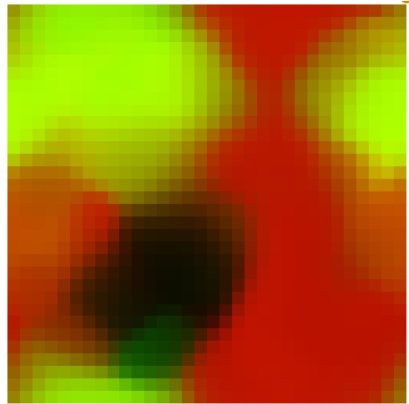
# Antigen profiles for the different entities



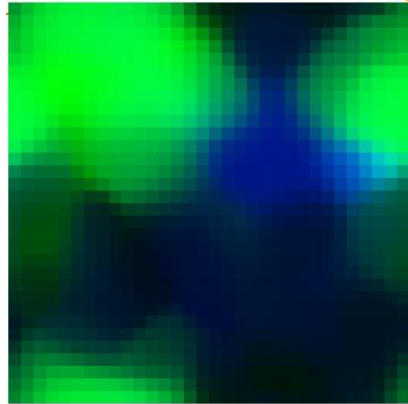
# Exemplary self-organizing maps (SOM)



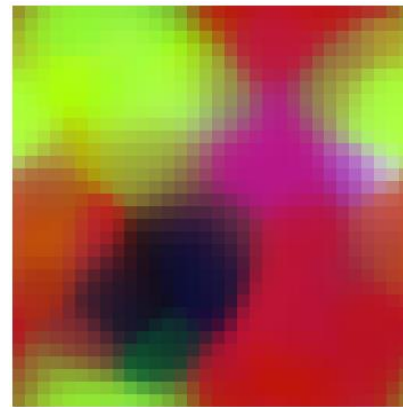
## Marker visualization



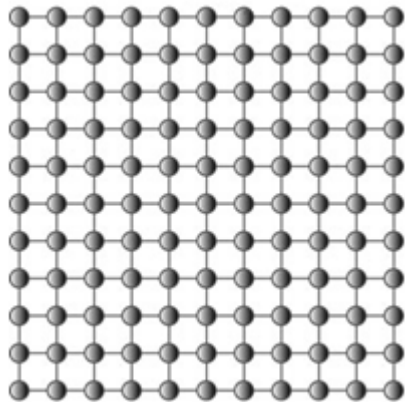
CD45 / SS



SS / CD19

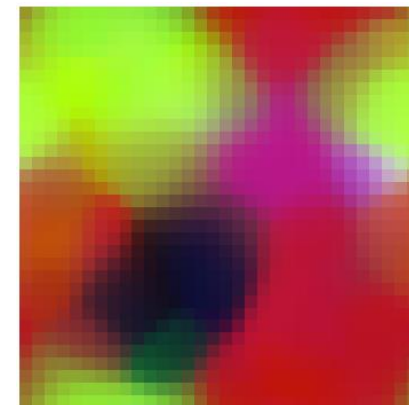


CD45 / SS / CD19  
CD45+ SS- CD19+

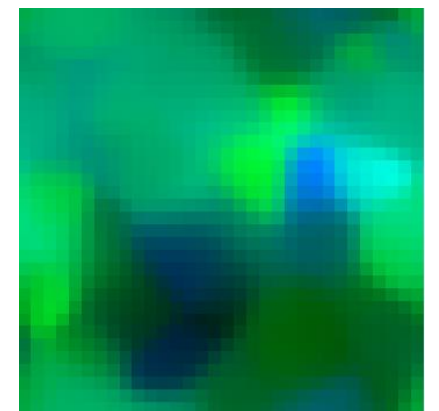


SOM  
32x32

## Kappa/Lambda representation



CD45 / SS / CD19  
CD45+ SS- CD19+

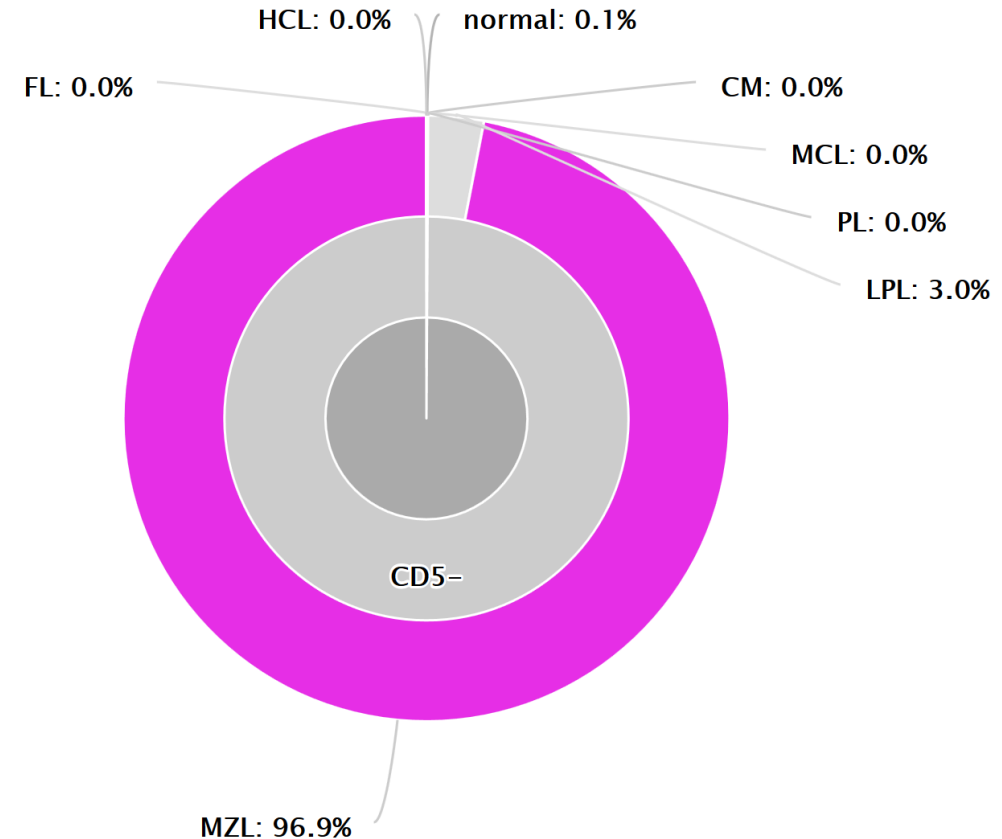
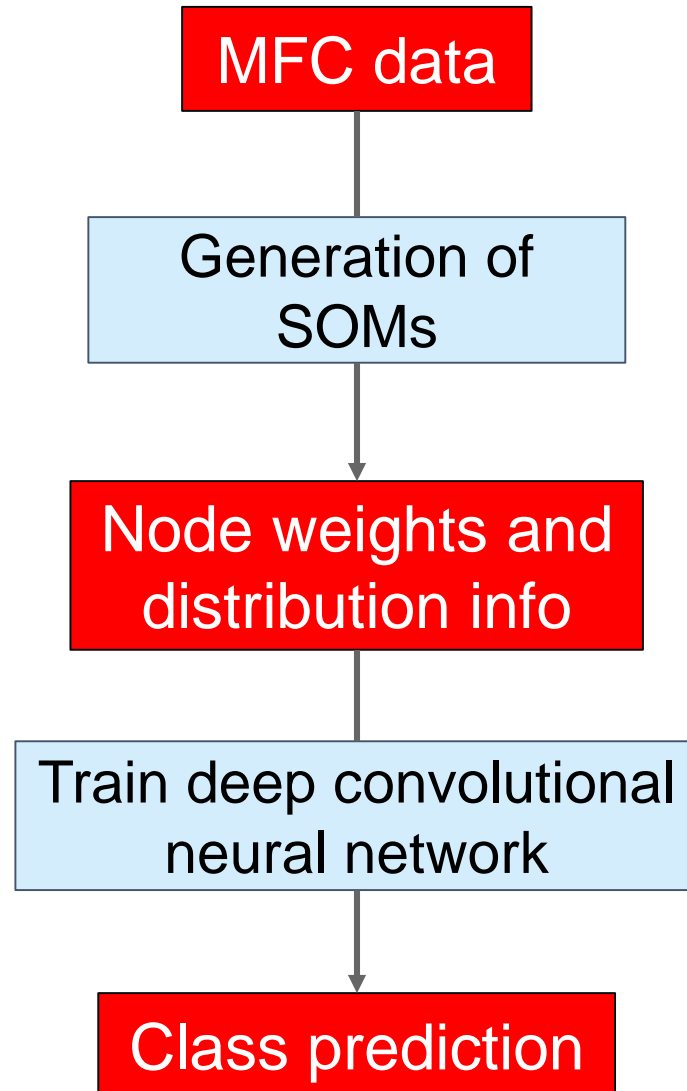


Kappa / Lambda

# AI approach for automated flow cytometric diagnosis



Workflow overview



→ Automated approach tested in parallel to expert routine diagnostics

# Study population



**3272 patients with  
suspected mature  
B-cell neoplasm**



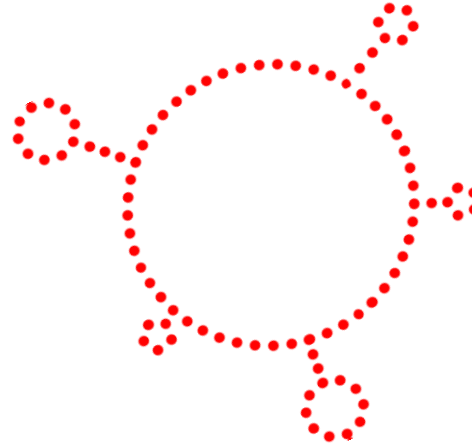
**n = 2304**



**n = 968**



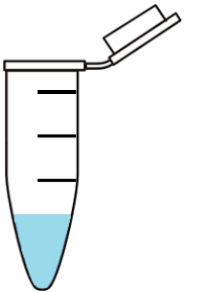
**April – July 2019**



**Flow cytometry based  
on two 9-color tubes**

**Tube 1:**

FMC7, CD10, IgM, CD79b, CD20,  
CD23, CD19, CD5, CD45



**Tube 2:**

Kappa, Lambda, CD38, CD25, CD11c,  
CD103, CD19, CD22, CD45



# Classes analyzed



Class	Absolute number of cases	Relative number of cases
Chronic lymphocytic leukemia	481	14.7%
CLL/PL	19	0.6%
Follicular lymphoma	16	0.5%
Hairy cell leukemia	61	1.9%
Variant hairy cell leukemia	3	0.1%
Lymphoplasmacytic lymphoma	46	1.4%
Mantle cell lymphoma	29	0.9%
Marginal zone lymphoma	11	0.3%
Monoclonal B-cell lymphocytosis	229	7.0%
No evidence of mature B-cell neoplasm	2377	72.6%



# Evaluation strategy

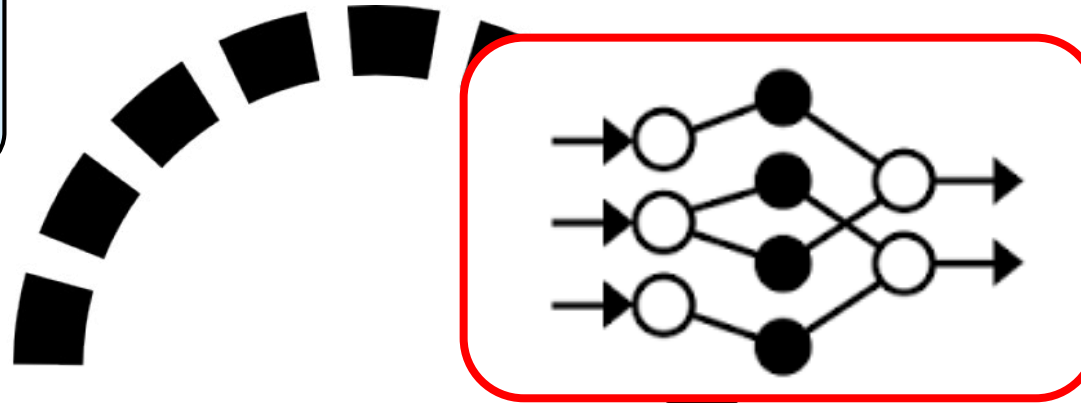


## Included

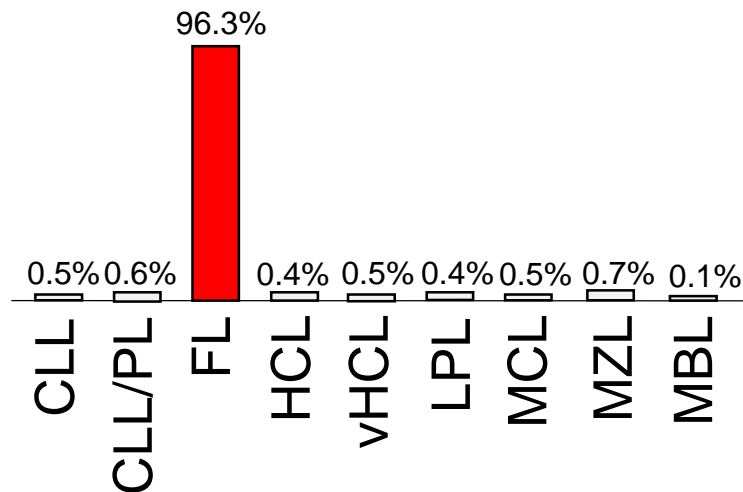
- 778 cases with infiltration rate  $\geq 1\%$
- 2377 negative cases

- 117 cases with infiltration rate  $< 1\%$

## Excluded



## Deep convolutional neural network



## Class probabilities

Cases with probabilities  $\geq 95\%$  were compared to expert analysis results (2445/3155)

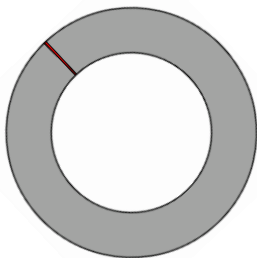
# Prediction accuracy

mature B-cell neoplasm vs. normal

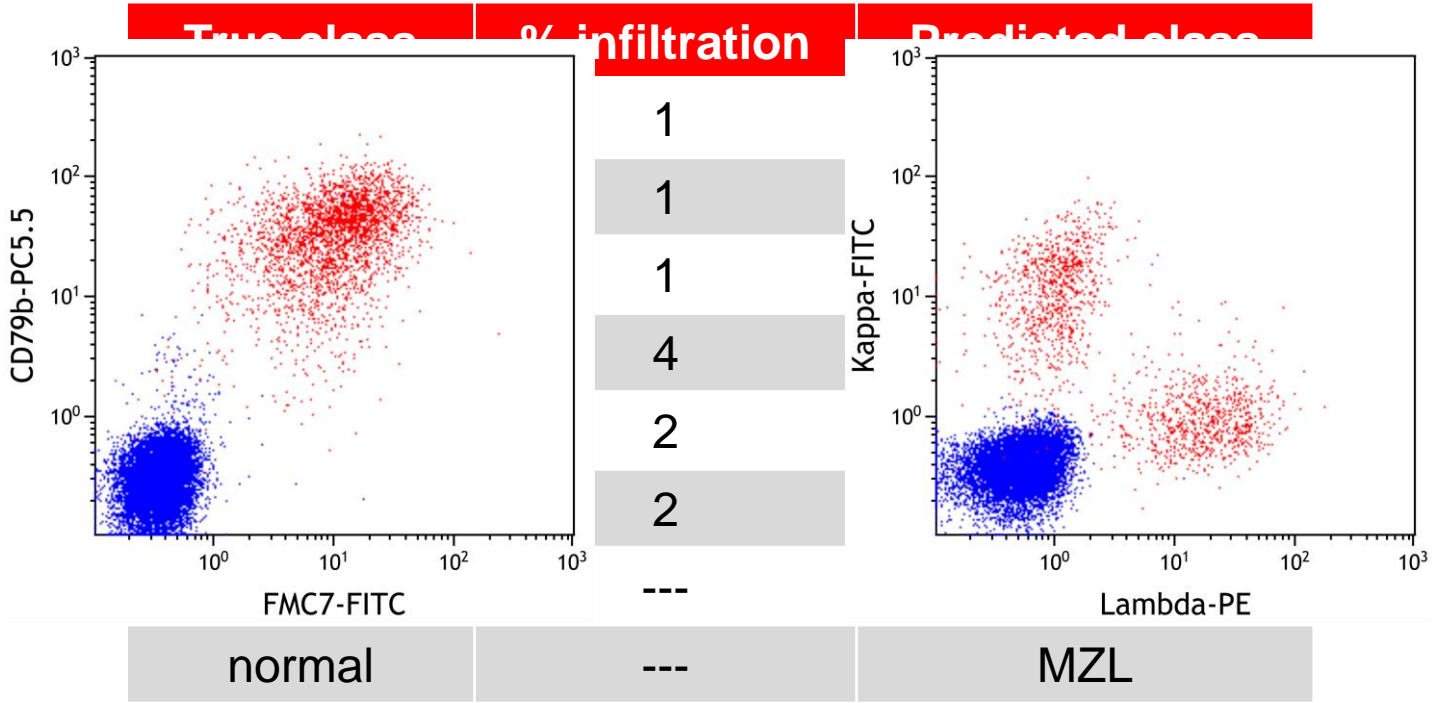


Prediction correct

2437/2445 (99.7%)



8 cases misclassified (3 BM, 5 PB)



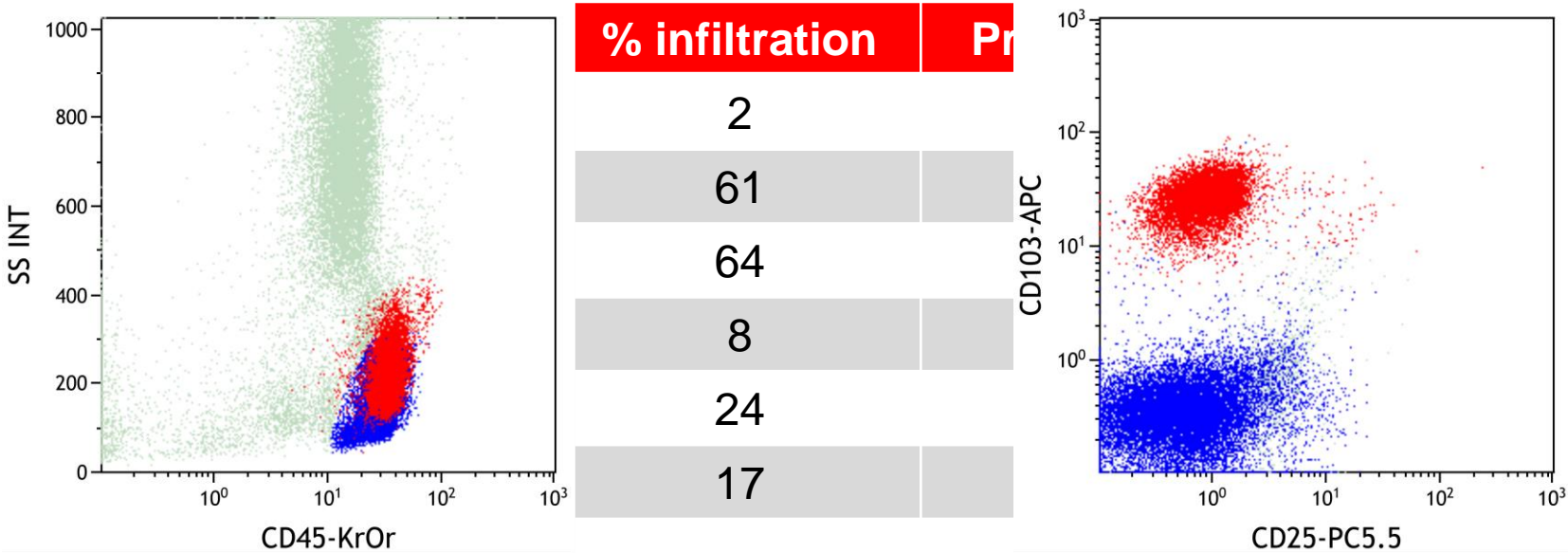
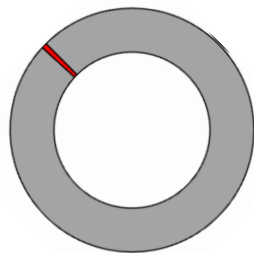
# Prediction accuracy

including 9 classes of mature B-cell neoplasm and normal

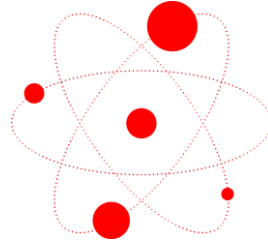


Prediction correct 2429/2445 (99.3%)

16 cases misclassified (5 BM, 11 PB)



# New matrix-based approach



**Amazon  
Web  
Services**

## Classifiers

- Decision tree
- Deep Learning model
- XGboost

## Sample categorization

- CLL, HCL, no lymphoma
- CD5+, HCL, other CD5-, no lymphoma
- Each lymphoma as its own class



6393 Samples



23 parameters



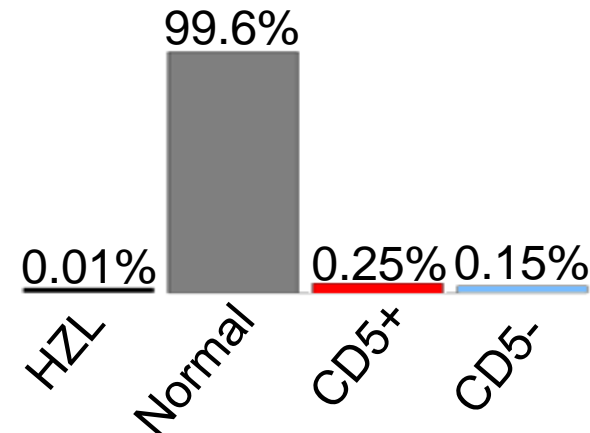
Validation



Training data



Test data



# Results: prediction accuracy

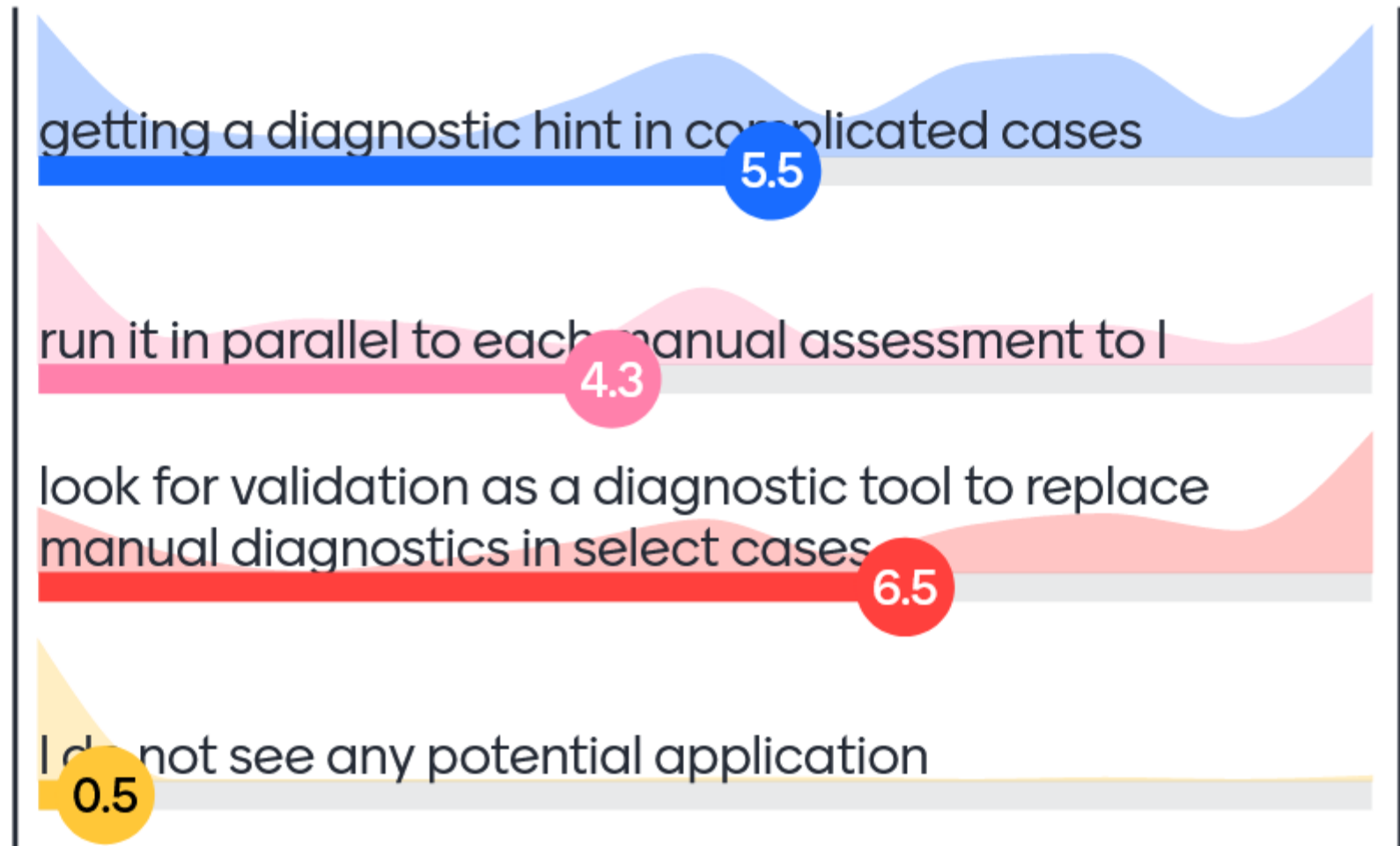


Sample categories	Selection of cases	Classifier, accuracy		
		DT	DL	XG
CLL, HCL, no lymphoma	>90% PP, clone size >0.1%	97%	99%	99%
CD5+, HCL, CD5-, no lymphoma	>95% PP, clone size >0.1%			96%
<b>Each lymphoma on its own</b>	<b>&gt;95% PP, clone size &gt;0.1%</b>			<b>93%</b>

- Continuous improvement and inclusion of other hematological entities
- BELUGA study



What would be your most likely application of AI/ML solutions for analysis of flow data?

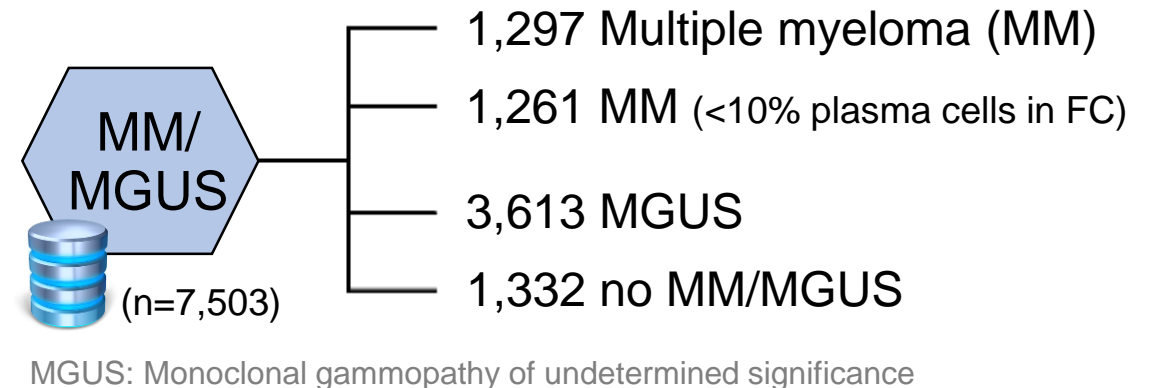
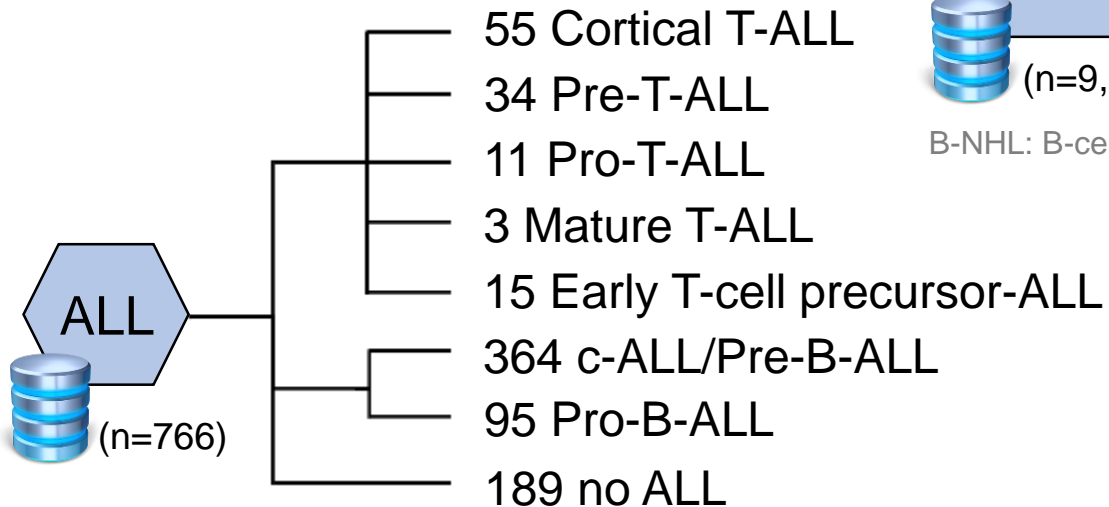
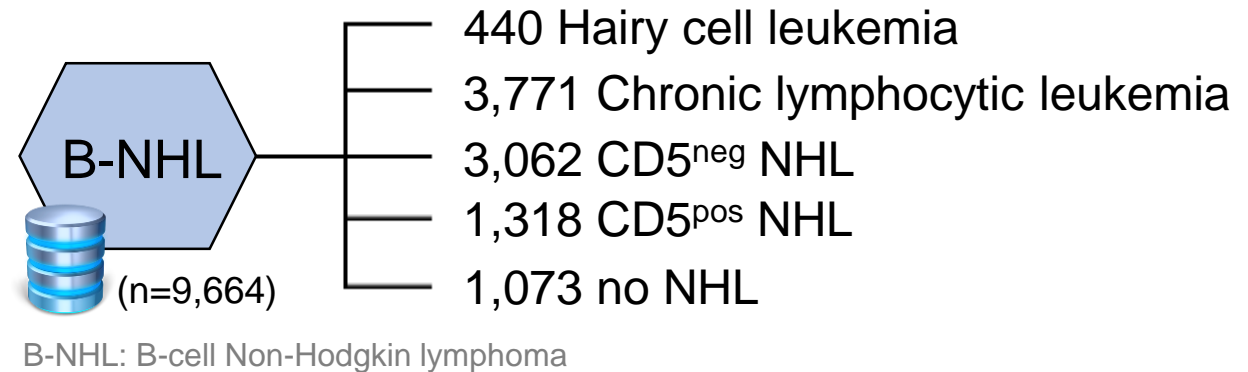
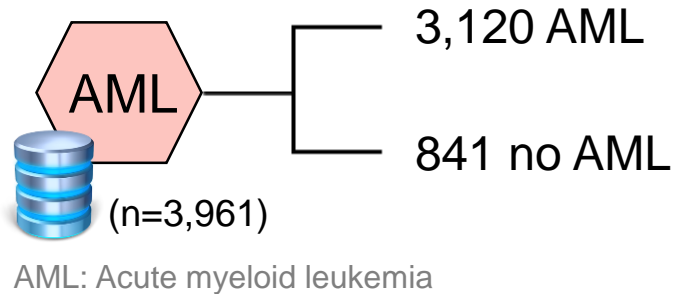
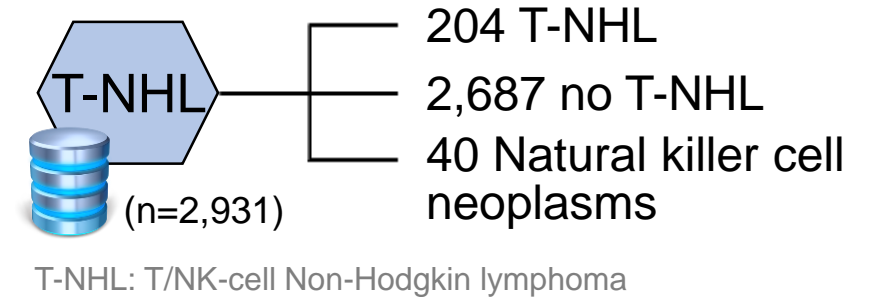
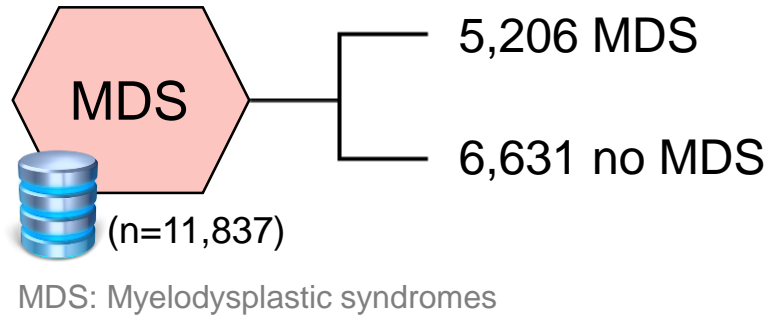


# Overview of the used dataset



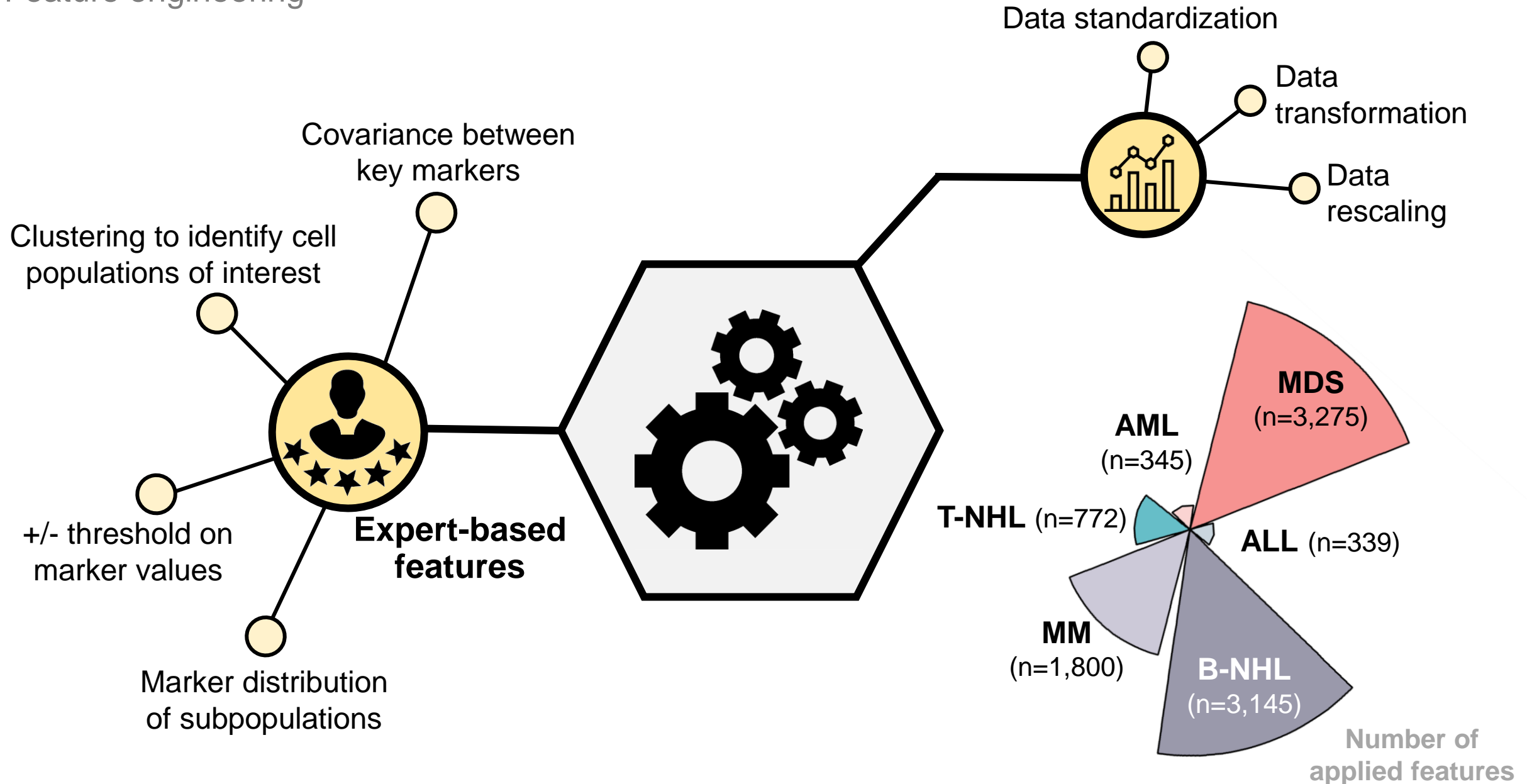
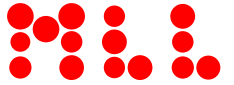
$\Sigma$  **36,662 cases**

- five-fold cross validation
- 80/20 training/test

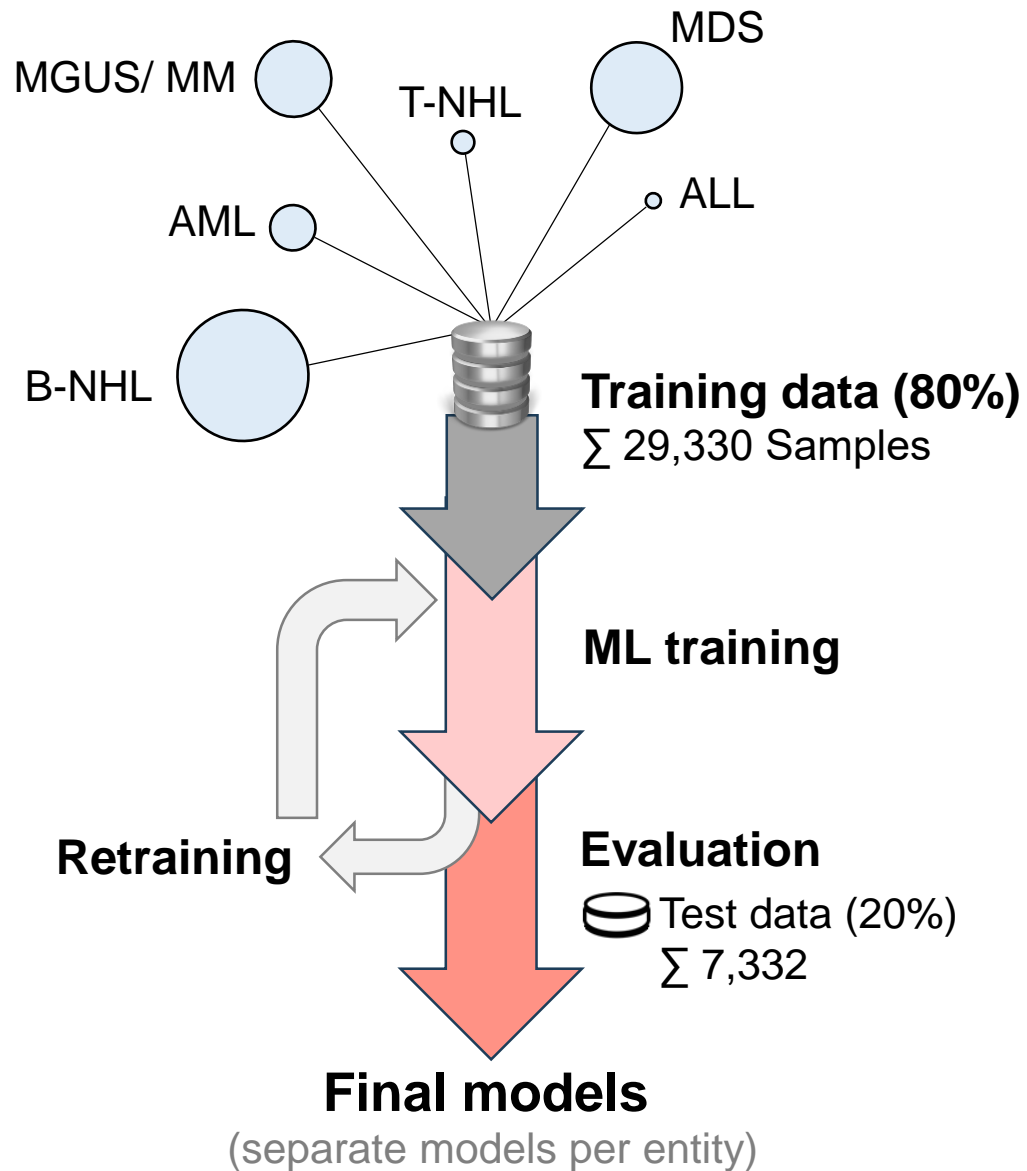


# Identification of relevant features

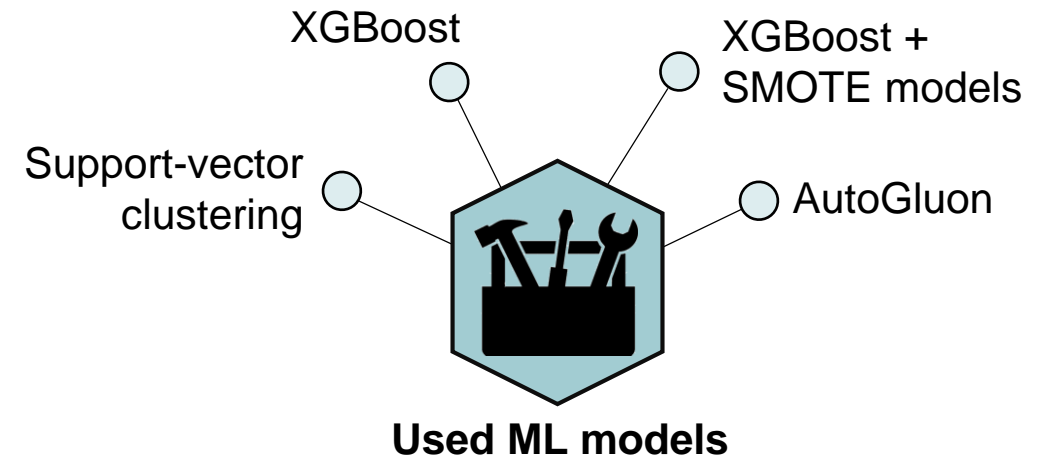
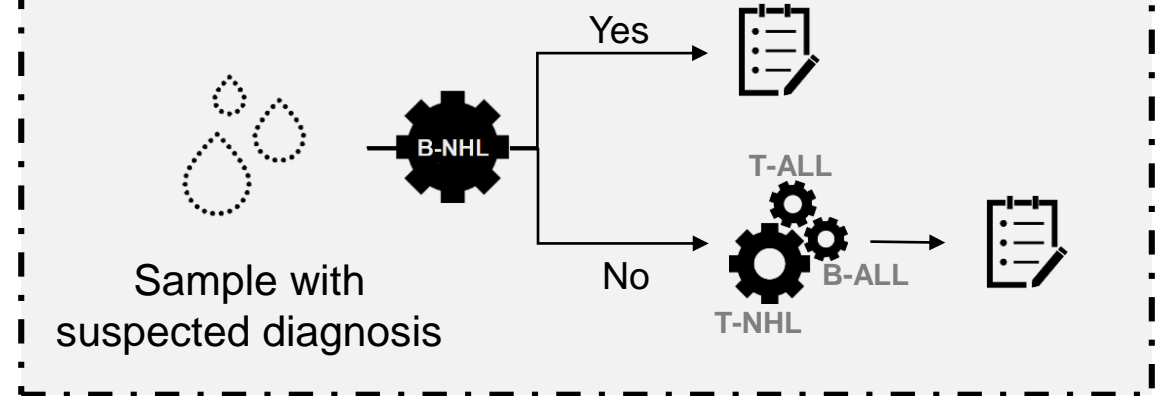
Feature engineering



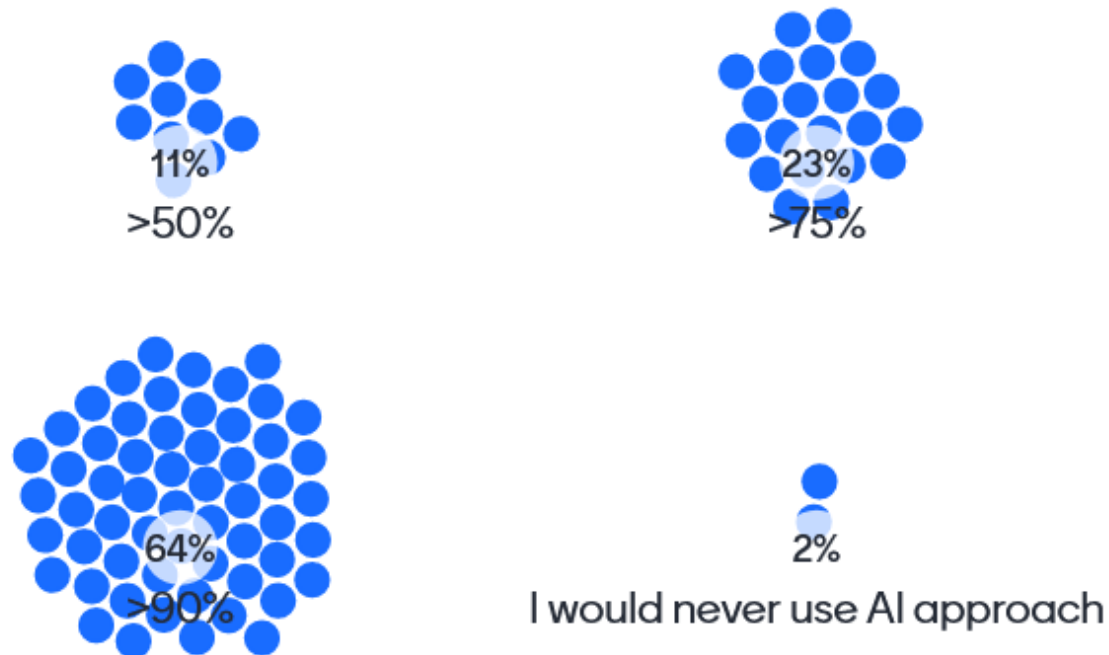
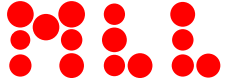
# Sample classification procedure



## Intended diagnostic workflow



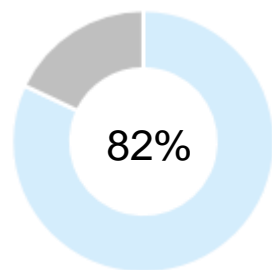
# Which portion of cases predicted at high prediction probability (PP) would you consider sufficient to decide using such AI approach?



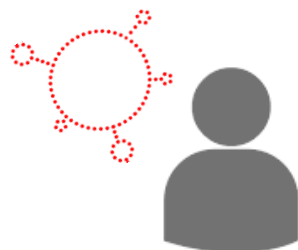


# Performance evaluation

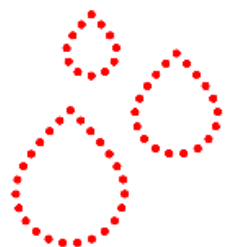
AML



% of cases with  
 $PP \geq 0.99$



XGBoost



AML?

Yes

**2,706**  
(100%)

2

No

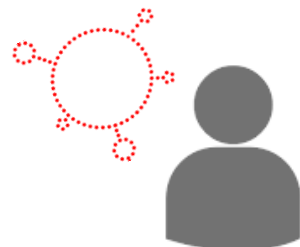
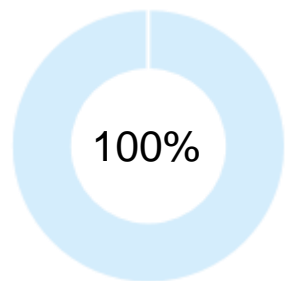
**552**  
(99.6%)

2

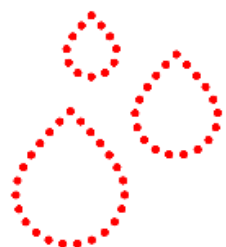
Average recall: 99.8%

# Performance evaluation

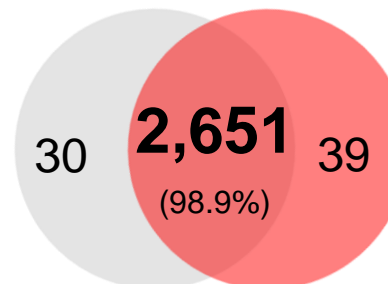
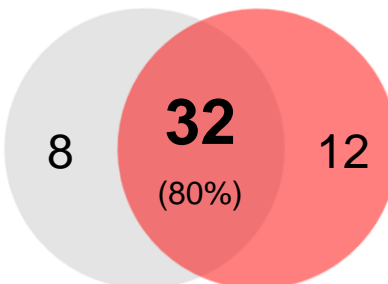
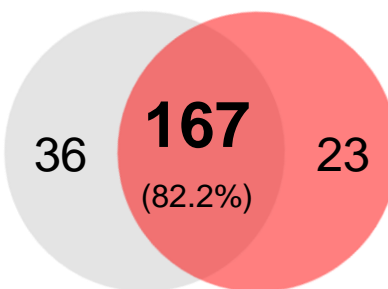
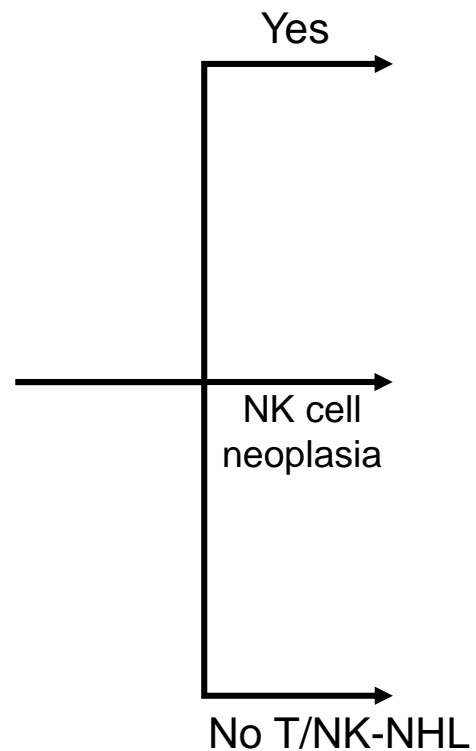
T-NHL



Weighted and linear SVC



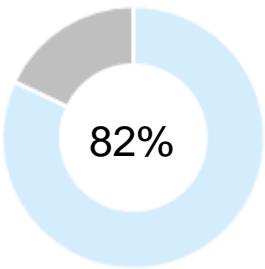
T-NHL?



Average recall: 87%

# Performance evaluation

ALL



% of cases with  
PP  $\geq$  0.9



XGBoost + SMOTE models

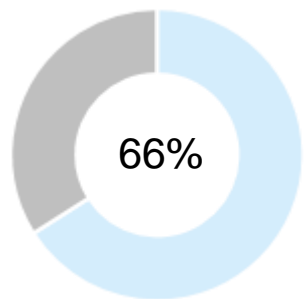


	Pro-B-ALL	c-ALL	Cortical T-ALL	non-cortical T-ALL	no ALL	
Pro-B-ALL	73	4	0	0	0	$\Sigma$ 77
c-ALL	6	319	0	0	5	$\Sigma$ 330
Cortical T-ALL	0	0	35	2	2	$\Sigma$ 39
non-cortical T-ALL	1	0	4	40	3	$\Sigma$ 48
no ALL	0	2	2	1	157	$\Sigma$ 162

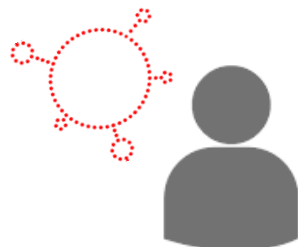
Average recall: 91.7%

# Performance evaluation

Multiple myeloma/MGUS

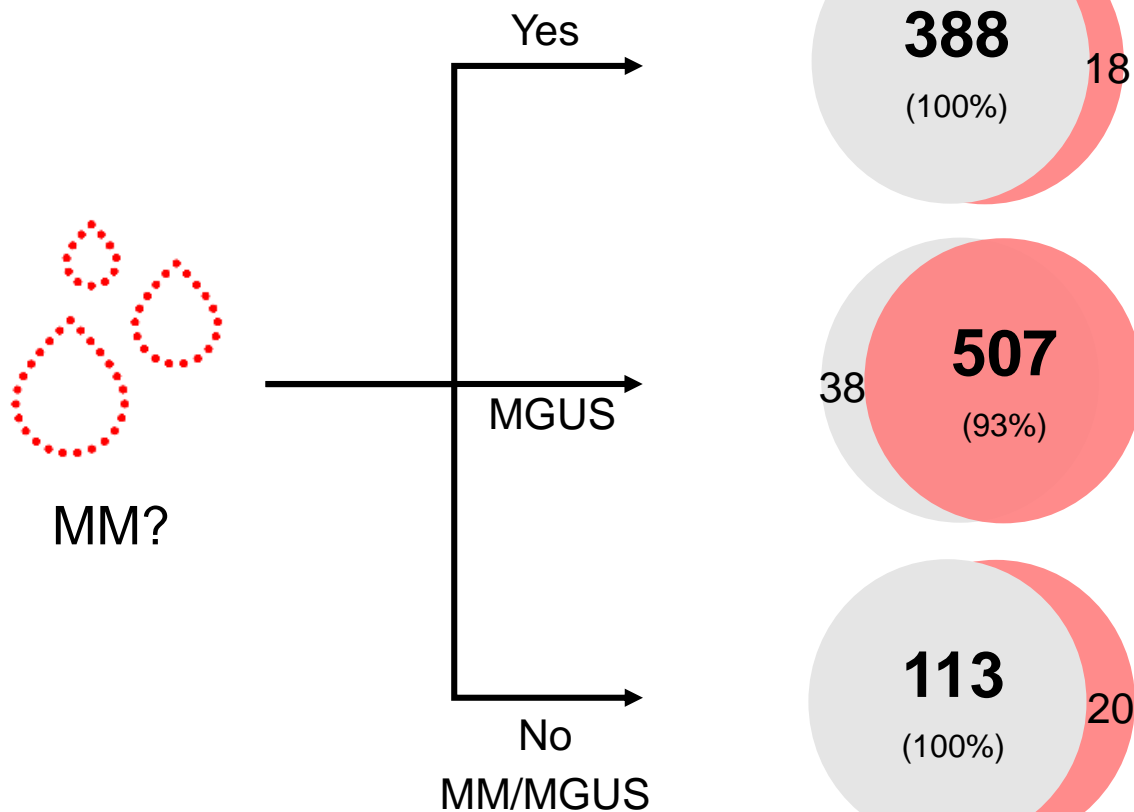


% of cases with  
 $PP \geq 0.9$



AutoGluon

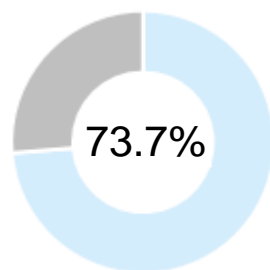
(weighted L2 ensemble of XGBoost, LightGBMXT & CatBoost)



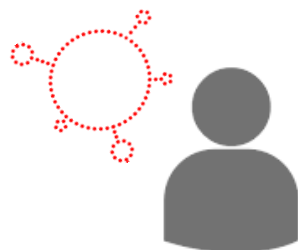
Average recall: 97.7%

# Performance evaluation

MDS

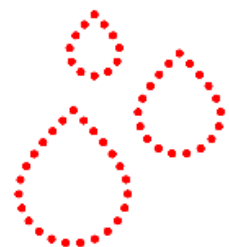


% of cases with  
 $PP \geq 0.9$



XGBoost

(+approach similar to manual gating strategies)



MDS?

Yes

82

**488**  
(85.6%)

128

No

128

**754**  
(85.5%)

82

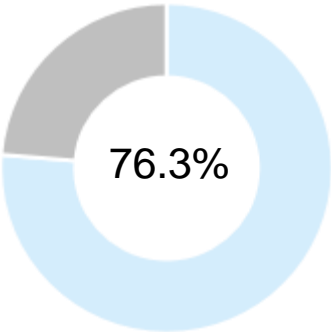
Average recall: 85.6%



# Performance evaluation



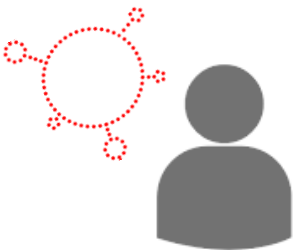
B-NHL



% of cases with  
 $PP \geq 0.92$



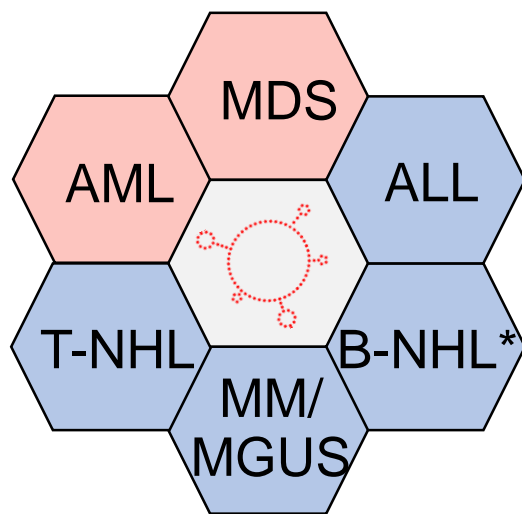
Hierarchical model



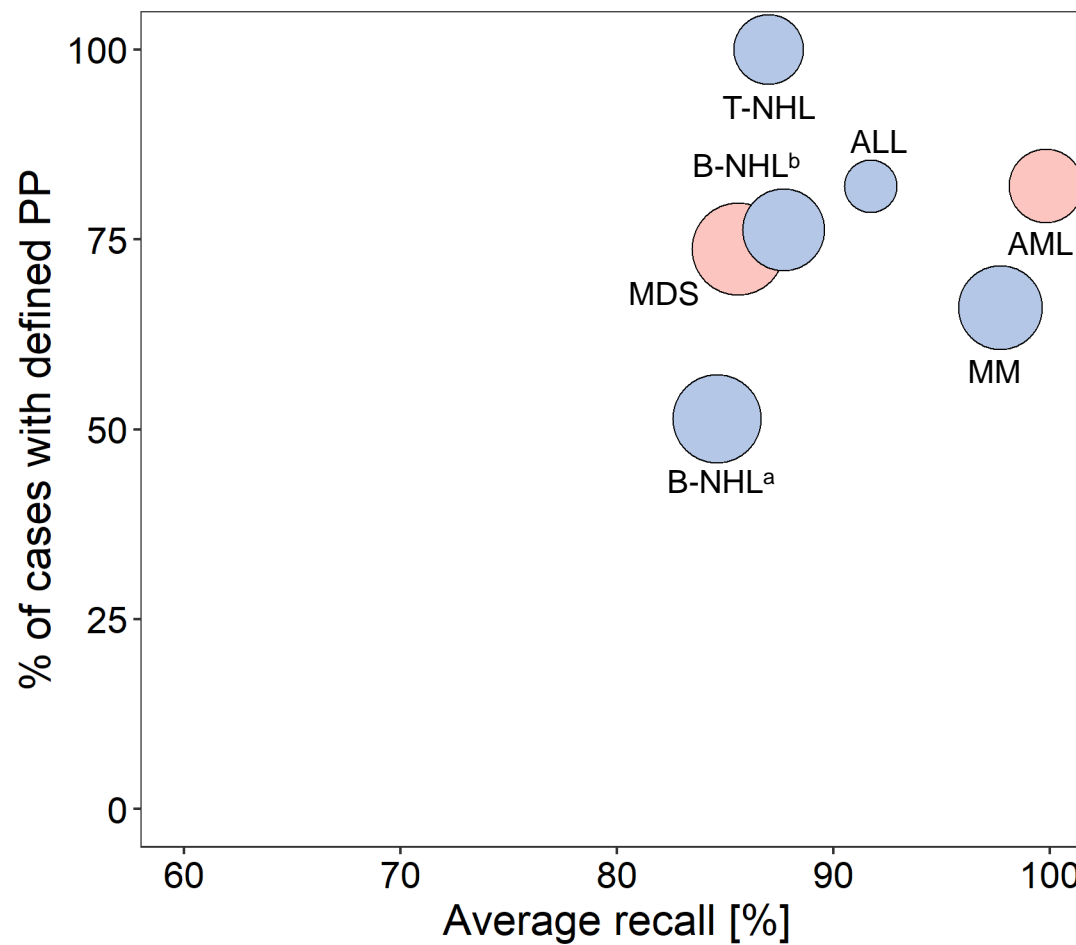
	CLL+CLL/PL+ MBL+MCL	HCL+HCL-v	LPL+MZL	no lymphoma	other	
CLL+CLL/PL+ MBL+MCL	3852	6	64	81	6	$\Sigma$ 4009
HCL+HCL-v	1	288	18	4	0	$\Sigma$ 311
LPL+MZL	54	3	1488	46	2	$\Sigma$ 1593
no lymphoma	1	1	4	754	0	$\Sigma$ 760
other	6	0	28	4	53	$\Sigma$ 91

Average recall: 87.7%

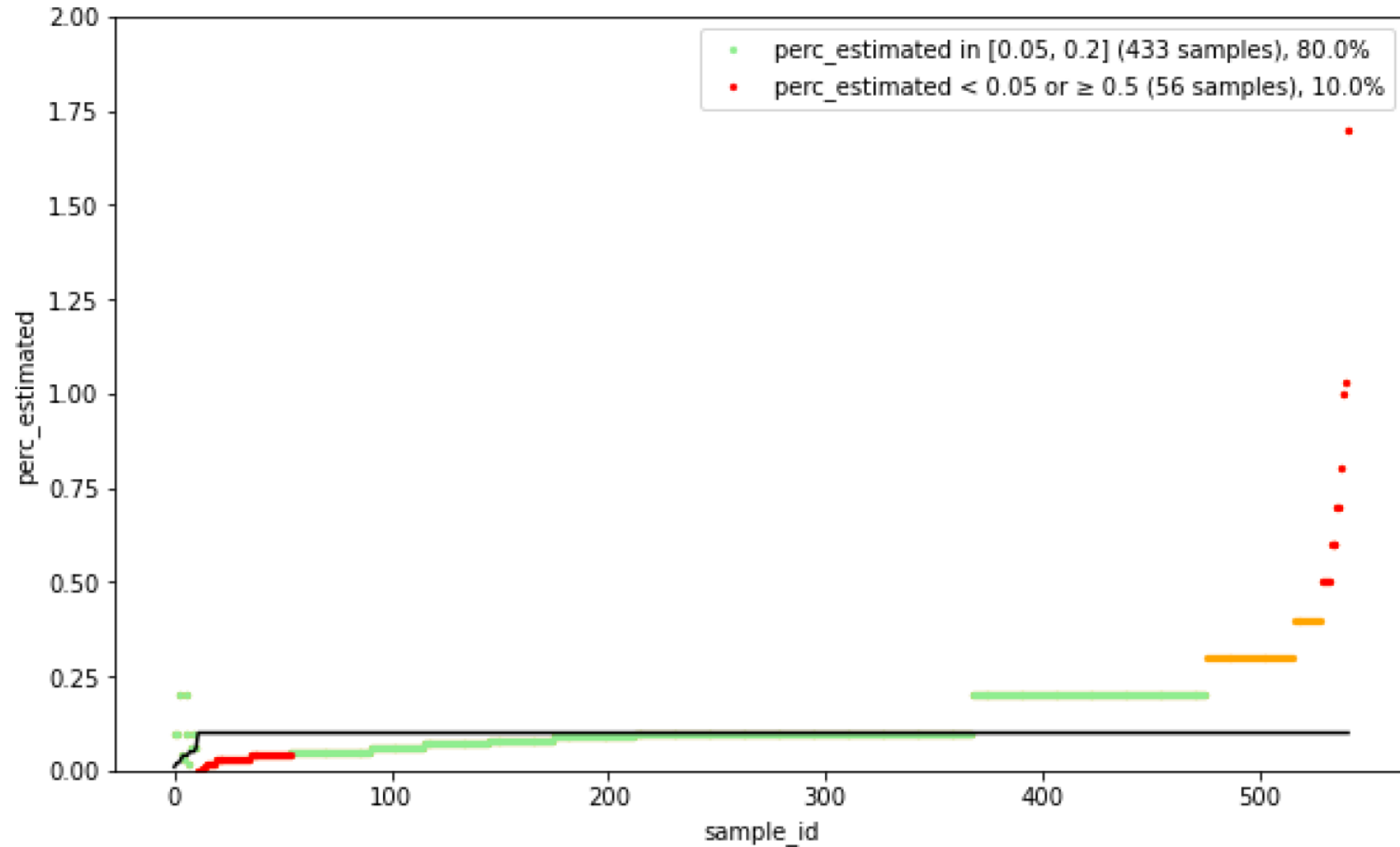
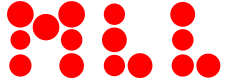
# Results summary



\*B-NHL<sup>a</sup> – first model  
B-NHL<sup>b</sup> – optimized model



# Application of AI for flow based MRD (MM)



# Automated classification of PB and BM cells



Metafer4 - MetaCyt

Mode File Slide Cells Scoring Training MetaCyt Configure Stage Filters Tools Help

Metafer 4

3.16.147

MetaCyt

33

1:1

Scan 2:

40x Leukocytes Classification/DNN

Scanning ...

Adm

Params : Blood-v8.DNN-Detection

Magnif : 39.8

Field : 0 of 79

Fields / s : 0.0

T Elapsed : 0:04

Remaining :

Setup

Search Fields

Gallery

Relocate

---

20-068009 TEST RD.1~A

0

???

???

???

???

20-068009 TEST RD.1~A

1

3


(Time lapse)

AI-based  
classification

MetaSystems Ikaros

[100%]

Datei Bearbeiten Ansicht Metaphase Filter Objekte Hilfe



12345mar

6789101112

131415161718

19202122X Y

Objektschwelle

Metaphase Maskieren




Objekte löschen




Objekte trennen




Überlappungen

Objekte prüfen

Beschriften







21-018349KE1-A ◀ 084a ▶ ◀ A ▶

1

2021-srv16 210309

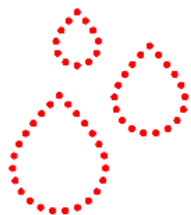
-870/-12512 CID:84

WP GBand





**WTS**  
(1.5d)



**DNA**  
(~6hr)

**WGS (~44hr)**  
90x coverage  
2x151 cycles



**WTS (~24hr)**  
50 mio reads  
2x101 cycles



**turn around time**  
**5-7 days**



Library  
Preparation

Sequencing

Data  
Analysis



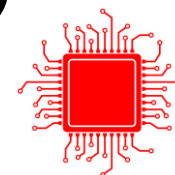
## Processing steps

- Fragmentation
- End repair
- Adapter Ligation
- Amplification (RNA)



## Data preprocessing (~7hr)

- FASTQ generation
- Alignment
- Variant calling (SV, SNV)

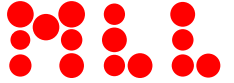


## Data analysis

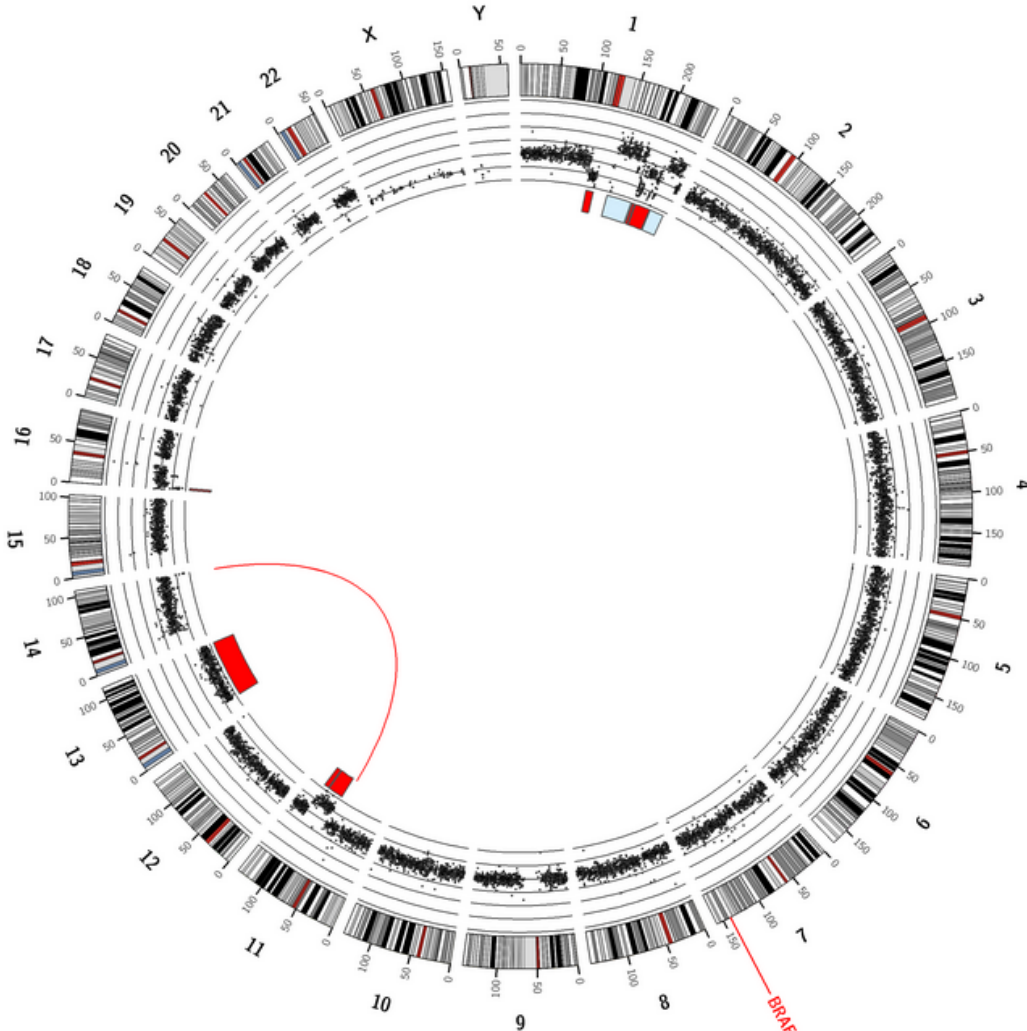
- Variant interpretation
- Gene expression
- CNV analysis
- SV analysis

*Bridger*

# Integrated application for WGS data



 Bridger   Select patient   Structural Variants ▾   Copy Number Variants   Small Nucleotide Variants ▾   Exon Coverage   RNA-seq   Final NGS report



## Copy Number Variants

- Gain
- Loss

## Structural Variants

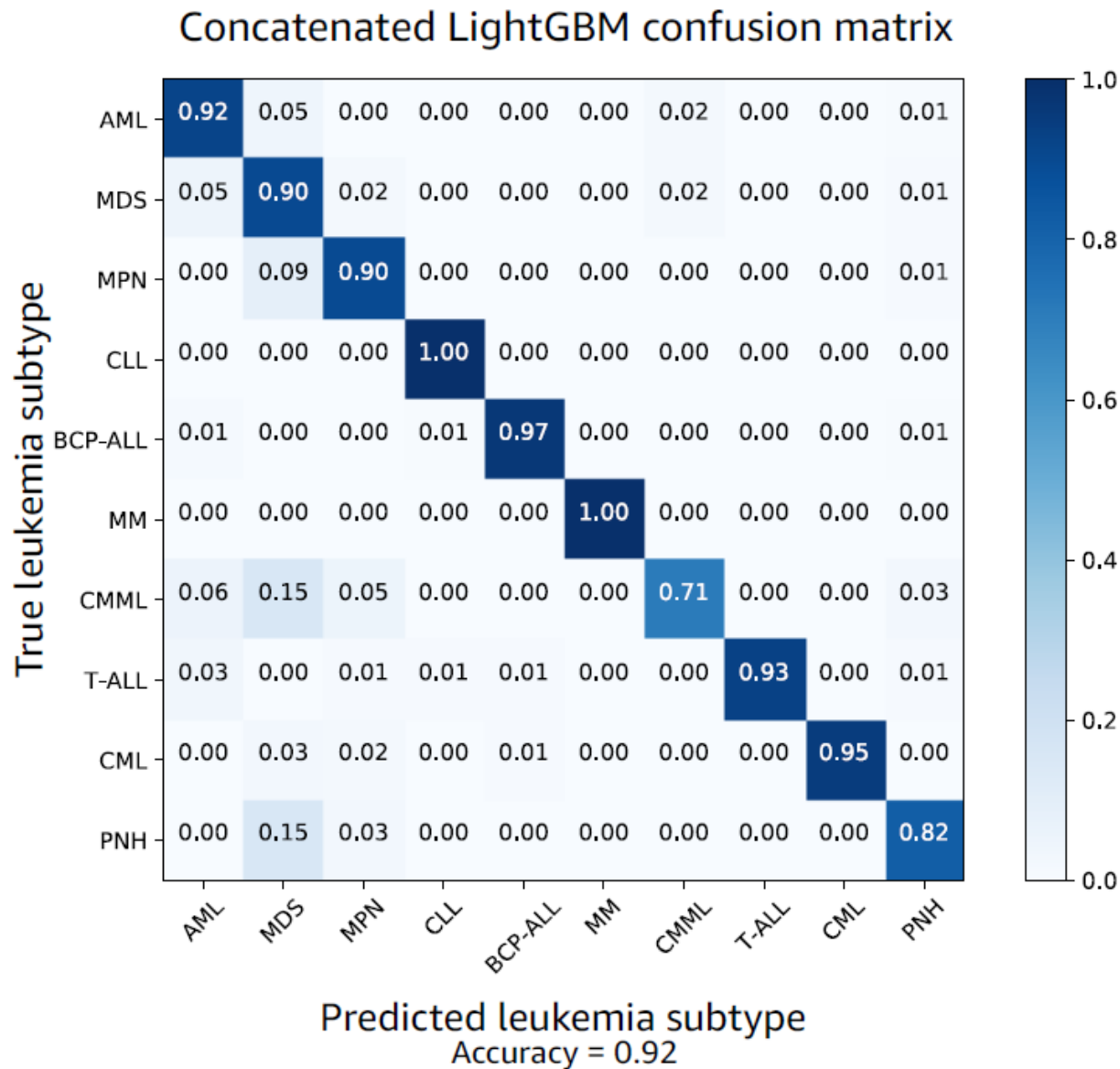
- Translocation





# MLL 5k genomes

## LGBM confusion matrix



# Integration diagnostics



- Implementation of AI and ML in diagnostic flow cytometry
- Implementation of AI and ML in flow based MRD assessment
- Implementation of AI and ML in cytomorphology, cytogenetics and NGS
- Application of WGS and WTS
- Integration of all methods by AI
- Classification according to WHO 2022

