Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF LIGHT

FAU

Master Thesis Presentation

Classification of cells using Real-Time Deformability Cytometry data

Ru Hui Tay

14th Oct 2022, 11:00

Agenda

Classification of cells using Real-Time Deformability Cytometry data

MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

01 Background

02 Problem Statement

03 Methods

04 Results

05 Conclusion

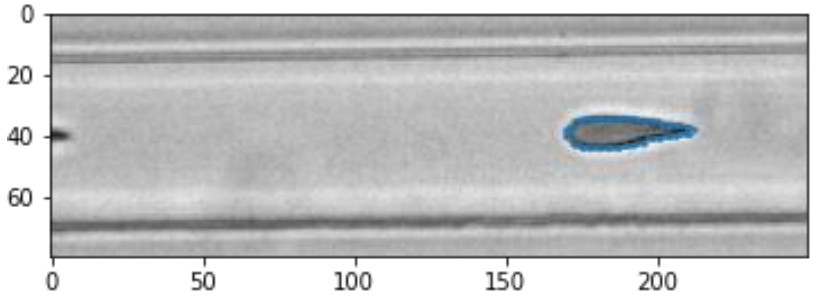
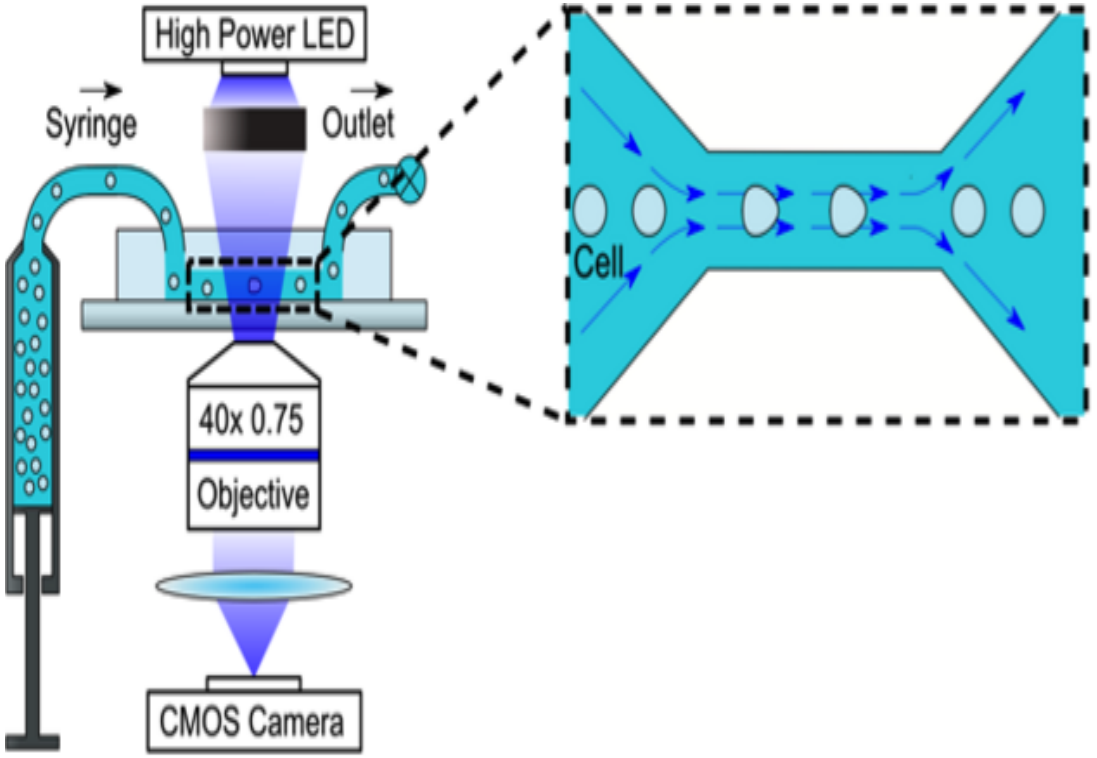


Image of a deformed singlet RBC taken by RTDC device. RBC is 14.62µm in length and 3.74µm in height.

Background

Background



Diagnosis

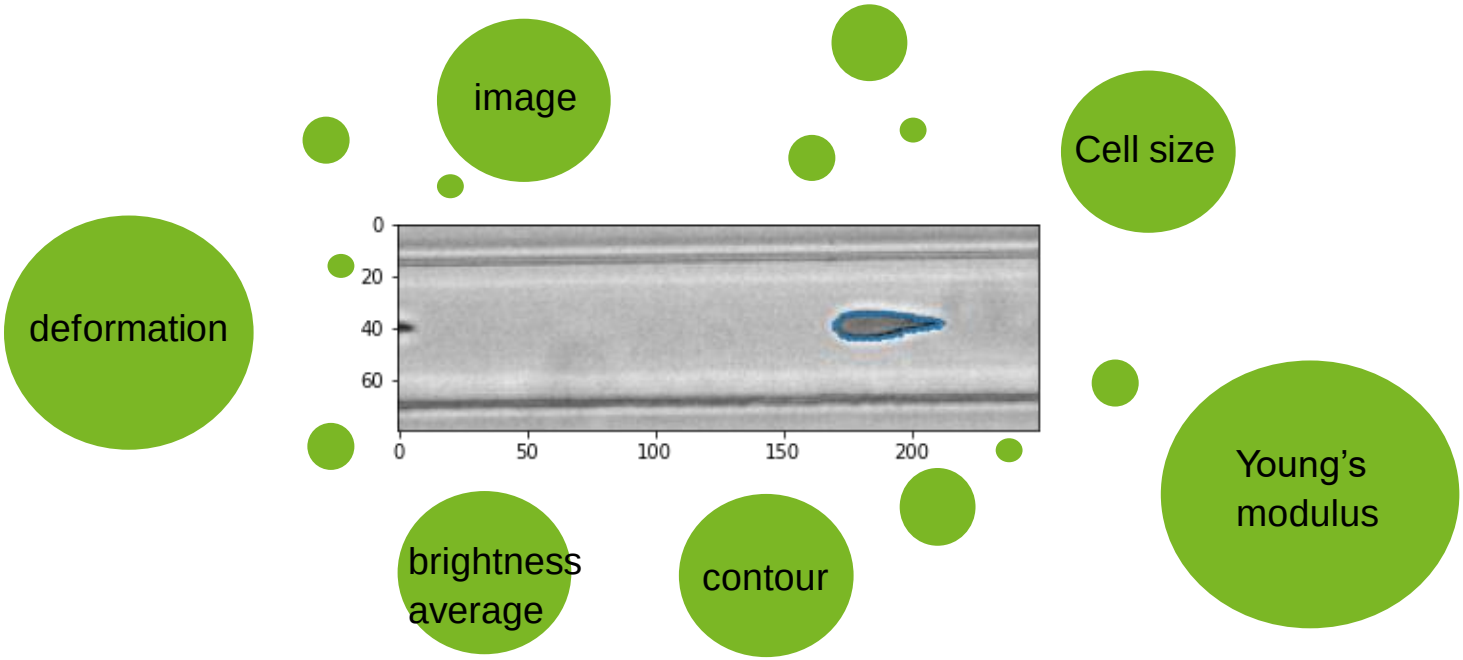
sick

not sick

https://www.google.com/url?sa=i&url=https%3A%2F%2Fcoronadashboard.government.nl%2Flandelijk%2Fziekenhuis-opnames&psig=AOvVaw2fBbFe1HscONzMyiuk_Oh5&ust=1665768950763000&source=images&cd=vfe&ved=0CAwQjRxqFwoTCMj5pIXf3foCFQAAAAAdAAAAABAE
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.legaltoday.com%2Fpractica-juridica%2Fderecho-penal%2Fpenal%2Futilizacion-en-el-proceso-penal-de-muestras-de-sangre-obtenidas-con-fines-terapeuticos-2014-11-21%2F&psig=AOvVaw1o2me32tXCH9n6b0lqoLHu&ust=1665769072340000&source=images&cd=vfe&ved=0CAwQjRxqFwoTCOiFpL_f3foCFQAAAAAdAAAAABAE
https://mpl.mpg.de/divisions/guck-division/methods/deformability-cytometry

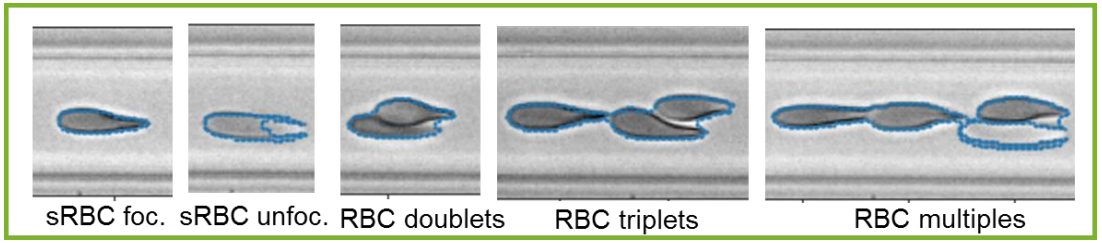
Common features the RTDC provides.

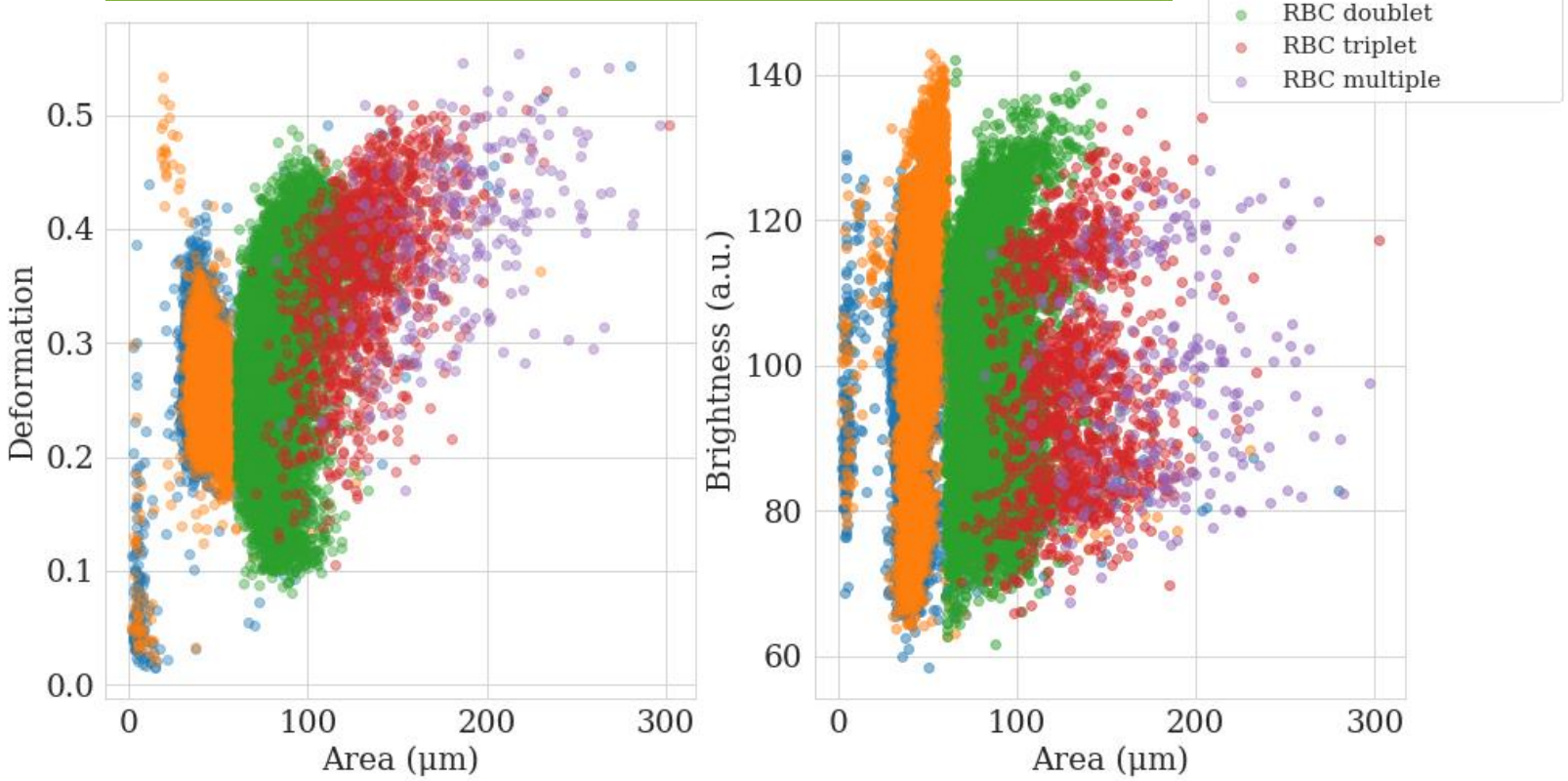
Background



RBC multiplets cannot be well-differentiated by scalar features.

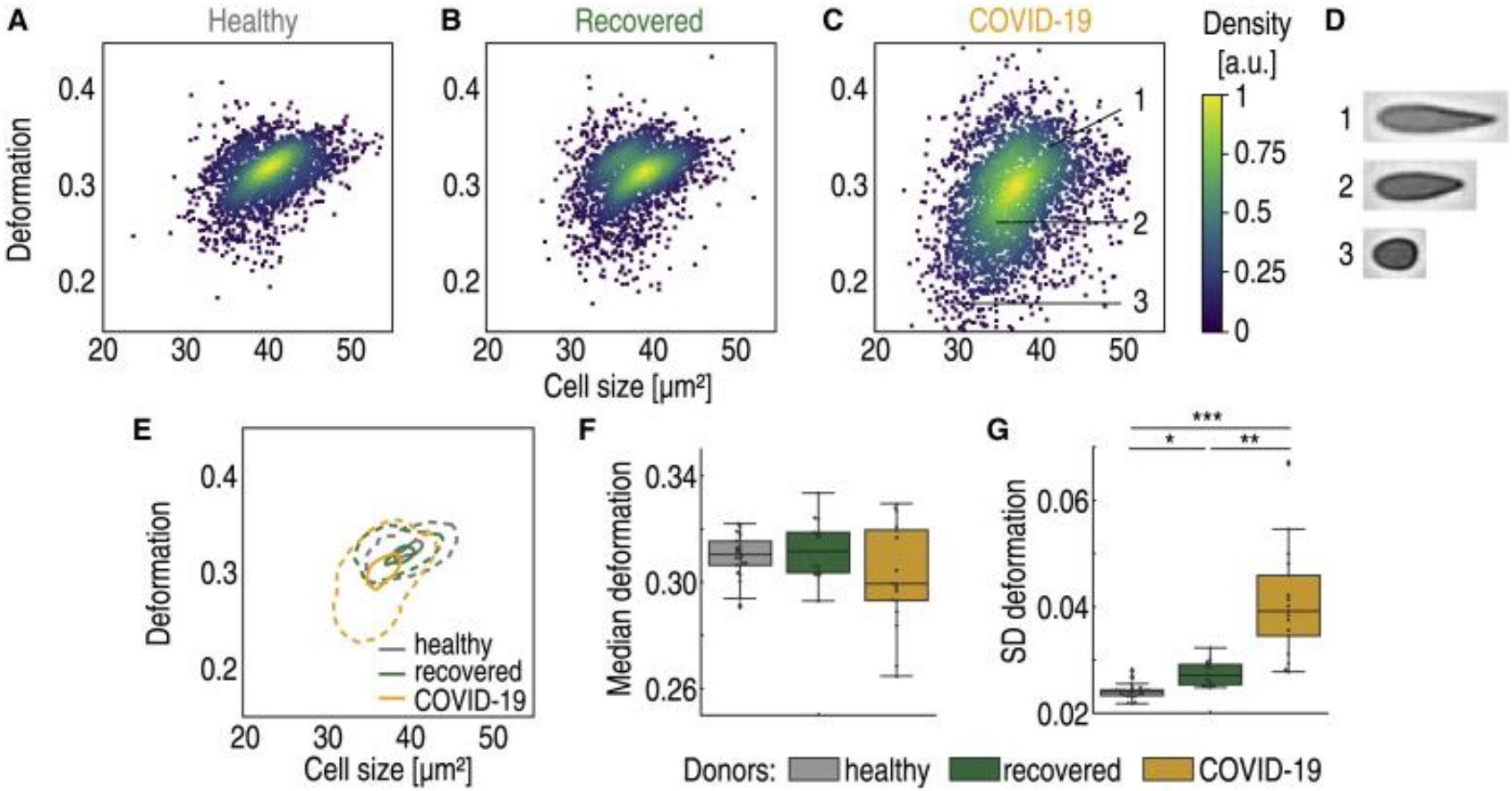
Background





Covid - singlet RBC differentiated by scalar features.

Background

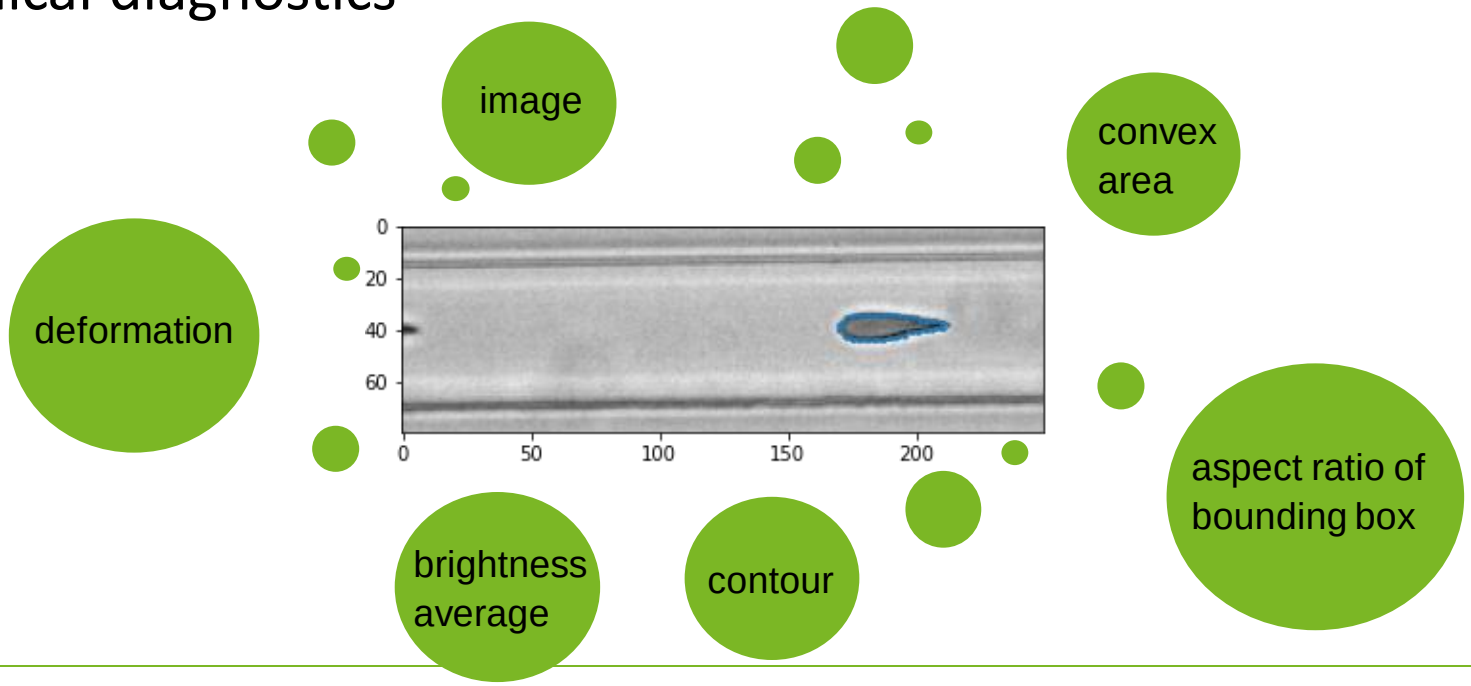


Kubánková, M., Hohberger, B., Hoffmanns, J., Fürst, J., Herrmann, M., Guck, J., & Kräter, M. (2021). Physical phenotype of blood cells is altered in COVID-19. *Biophysical journal*, *120*(14), 2838-2847.

Problem Statement

Problem Statement

- Scalar quantities are insufficient when cell types need to be distinguished
- Need a method to distinguish cell types → application in clinical diagnostics



Defining the boundaries of the thesis.

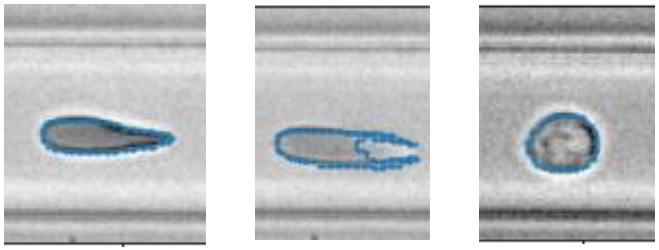
Problem Statement

MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

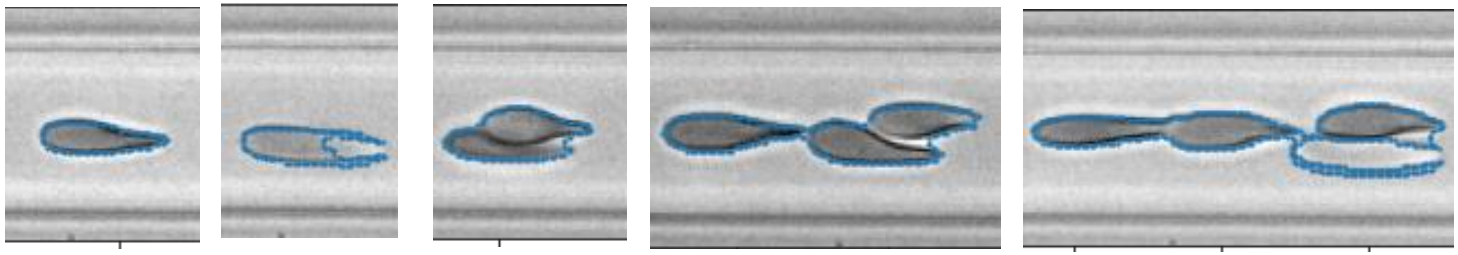
FAU

experiments

1



2



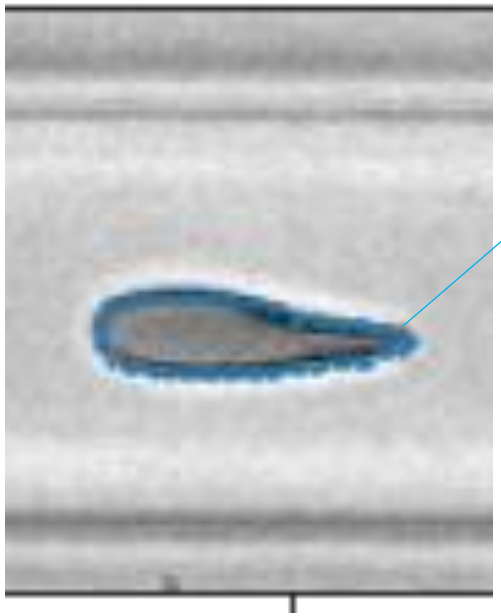
3

Classify between control versus long covid.

Tested 15 blood cell types.

How to classify contours?

Problem Statement

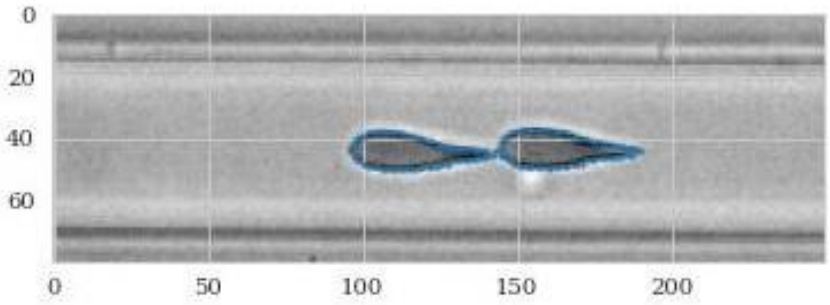


$$(X_1, Y_1)$$
$$(X_2, Y_2)$$
$$(X_3, Y_3)$$
$$(X_4, Y_4)$$
$$(X_5, Y_5)$$
$$\vdots$$
$$(X_K, Y_K)$$

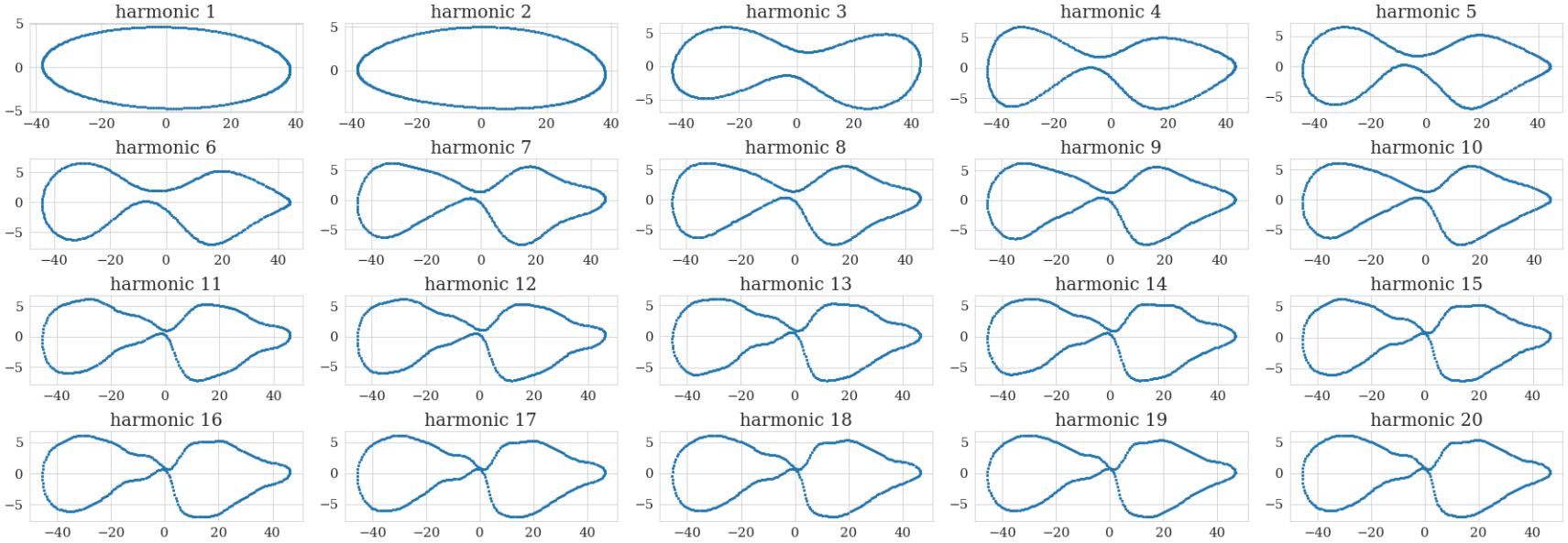
Methods

Fourier Descriptors

Methods



RBC doublet

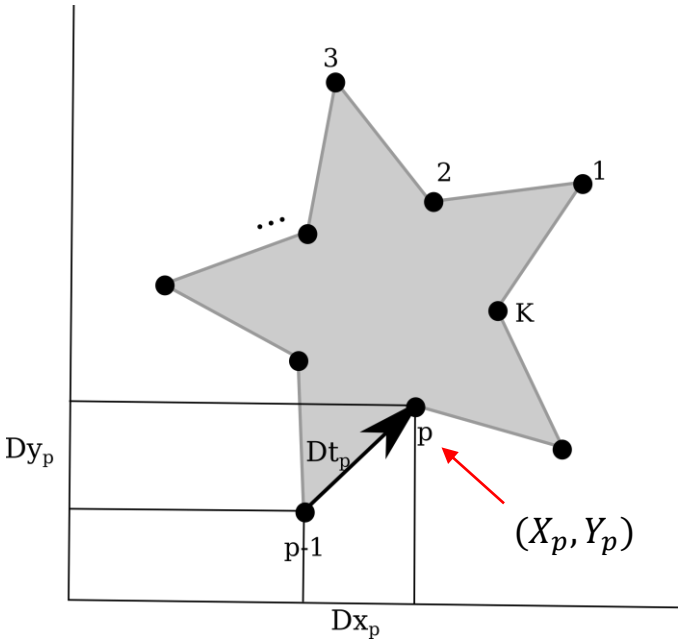


Fourier Descriptors

Methods

$$X_p = X_{cen} + \sum_{n=1}^{N} a_n \cos\frac{2n\pi t_p}{T} + b_n \sin\frac{2n\pi t_p}{T}$$

$$Y_p = Y_{cen} + \sum_{n=1}^{N} c_n \cos\frac{2n\pi t_p}{T} + d_n \sin\frac{2n\pi t_p}{T}$$

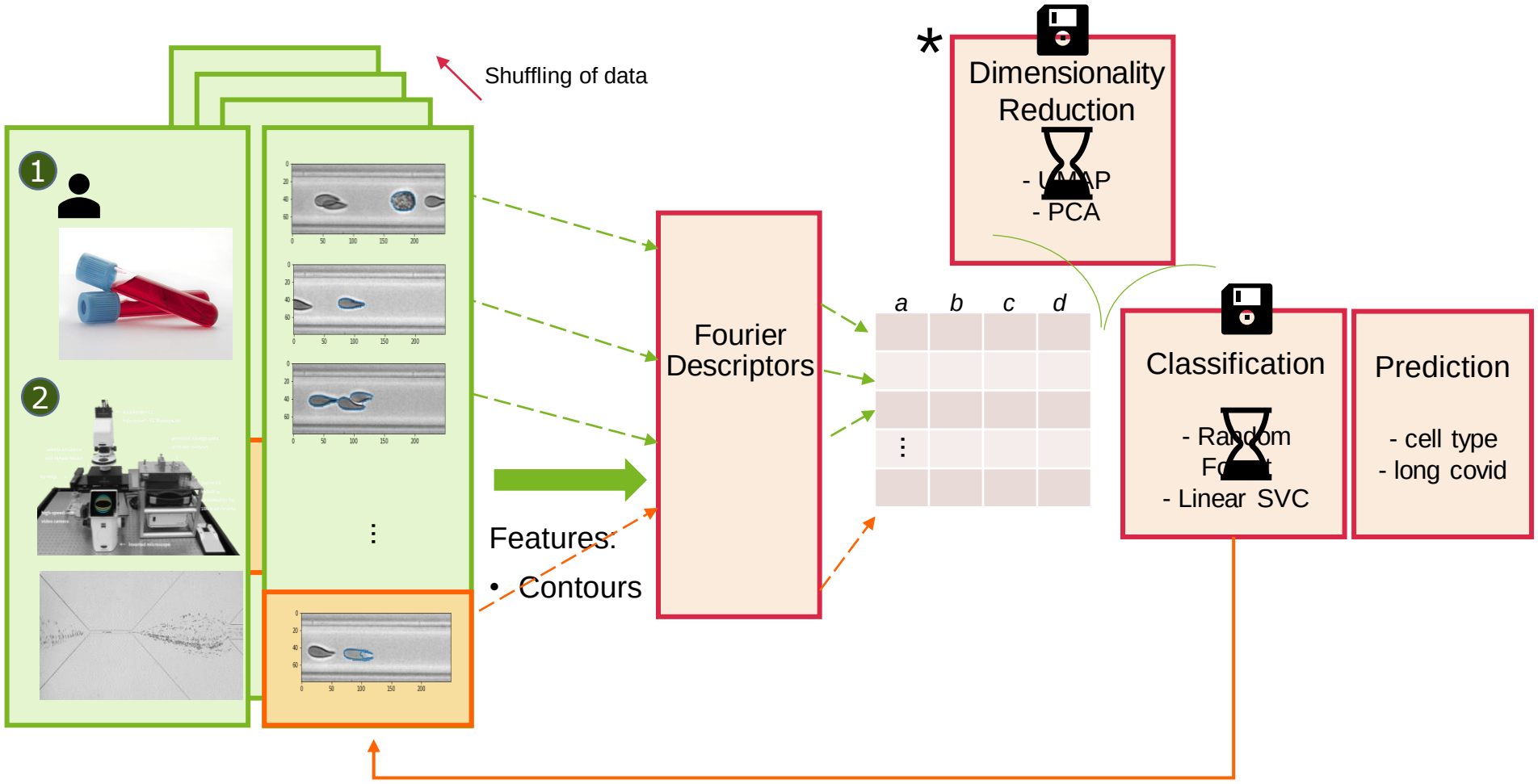


- (X_p, Y_p) – an arbitrary set of cartesian coordinates
- N – chosen harmonic number
- a_n, b_n, c_n, d_n - Fourier Descriptors coefficients
- $t_p = \sum_{j=1}^{p} Dt_j$
- $T = t_K$

Diaz, G., Zuccarelli, A., Pelligra, I., & Ghiani, A. (1989). Elliptic Fourier analysis of cell and nuclear shapes. *Computers and biomedical research*, *22*(5), 405-414

Analysis Pipeline

Methods



Results

Results

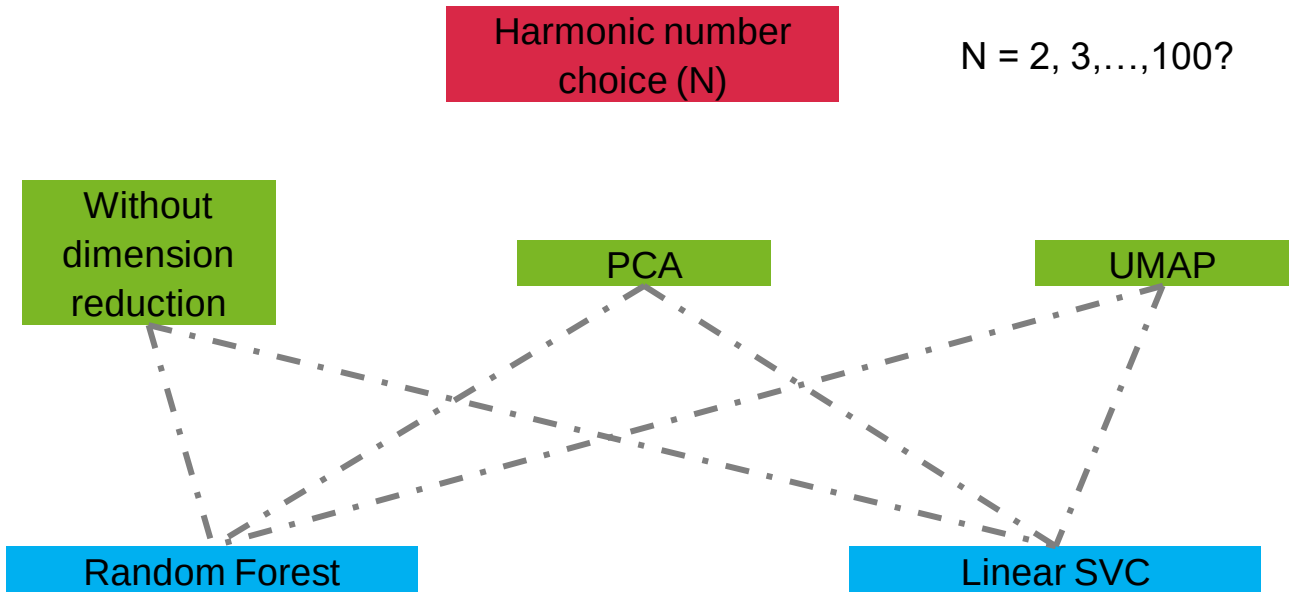
- Problem Statement: Can we classify cells using contours?
- Which combination of methods are best for classification?

Stage of the pipeline

Fourier Descriptor

Dimension reduction method

Classification method



Recap

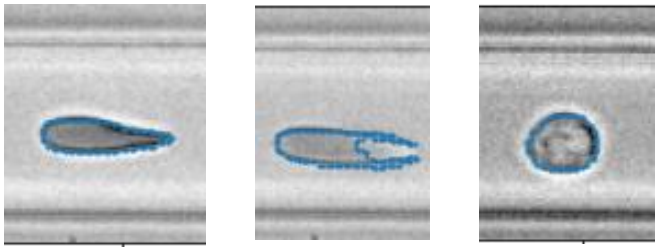
Defining the boundaries of the thesis.

MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

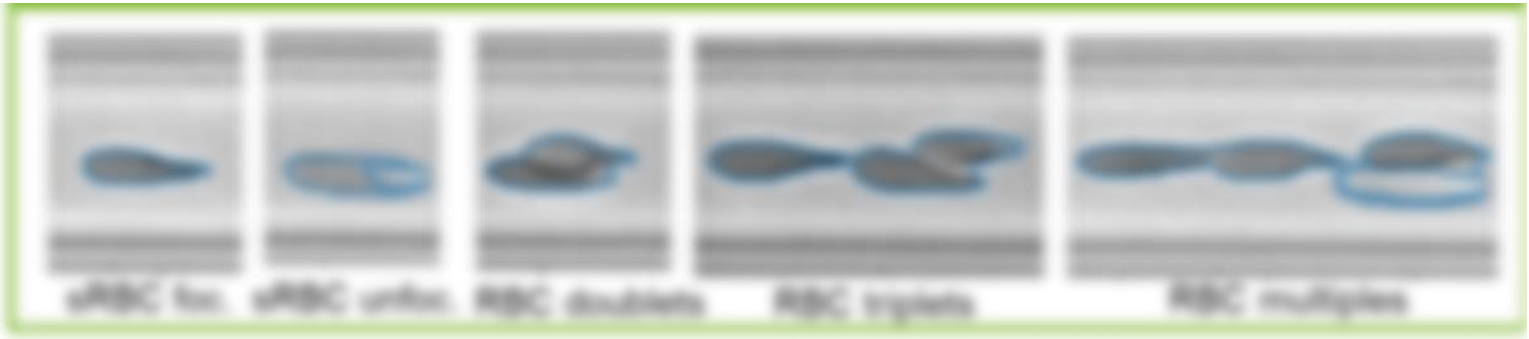
FAU

experiments

1



2



3

Classify between control versus long covid.

Tested 15 blood cell types.

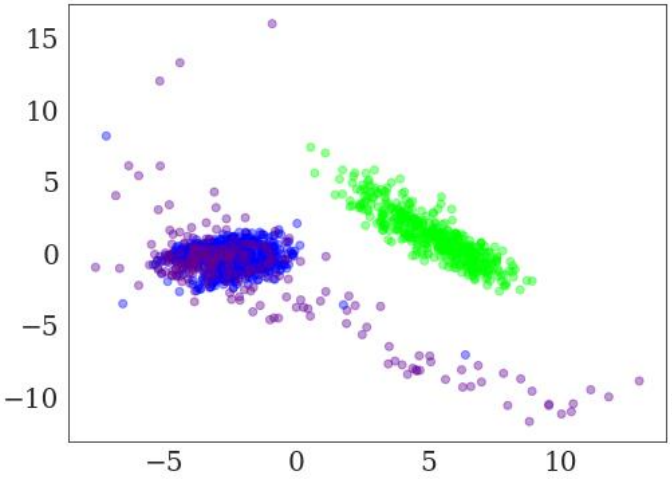
PCA (8→2 dimensions)

Experiment 1

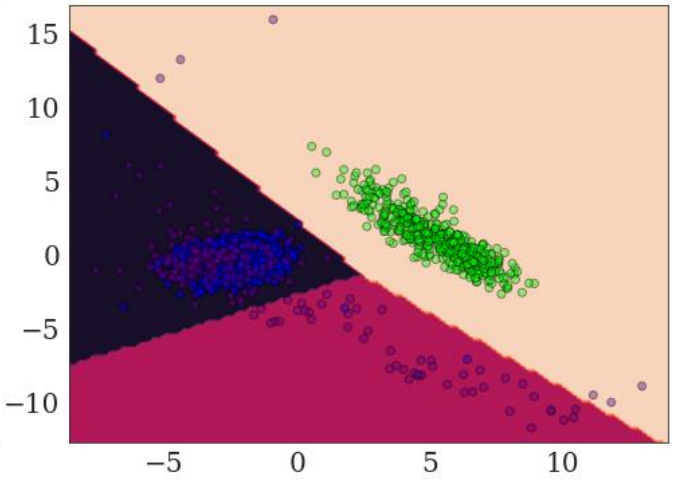
MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

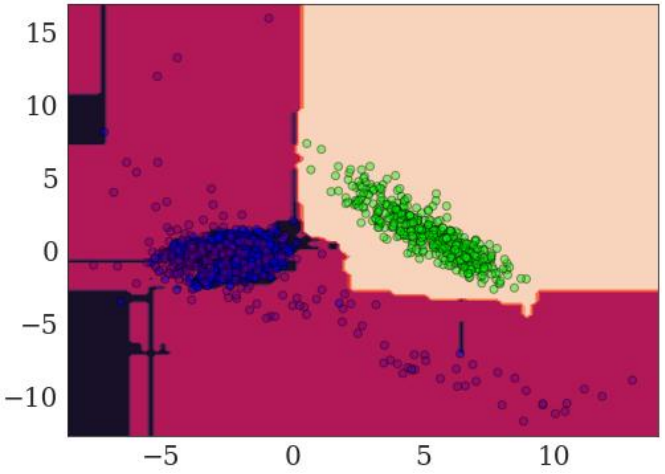
- Reduced from harmonic number 2
- PCA
- Random Forest - more sophisticated decision boundaries.



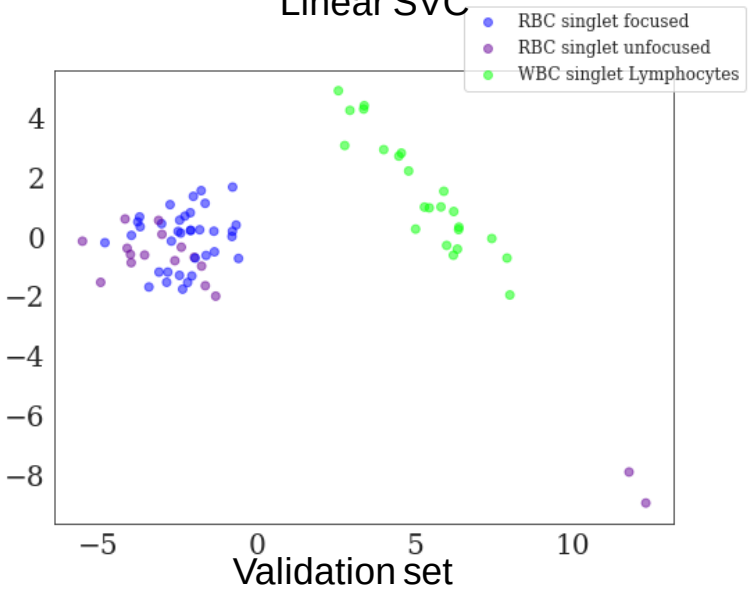
Training set



Linear SVC



Random Forest

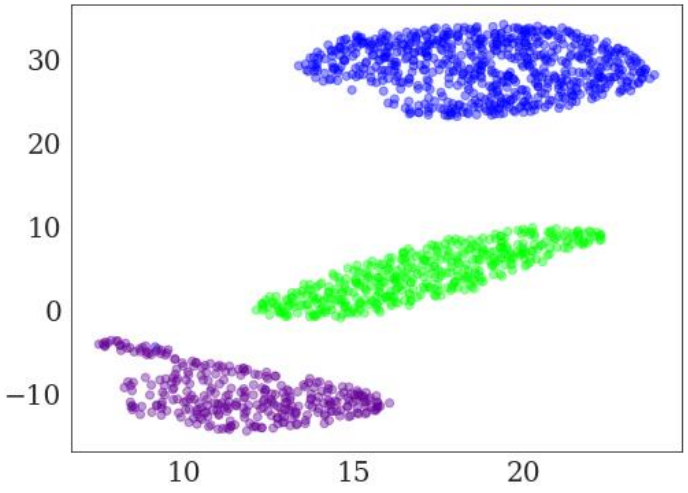


Validation set

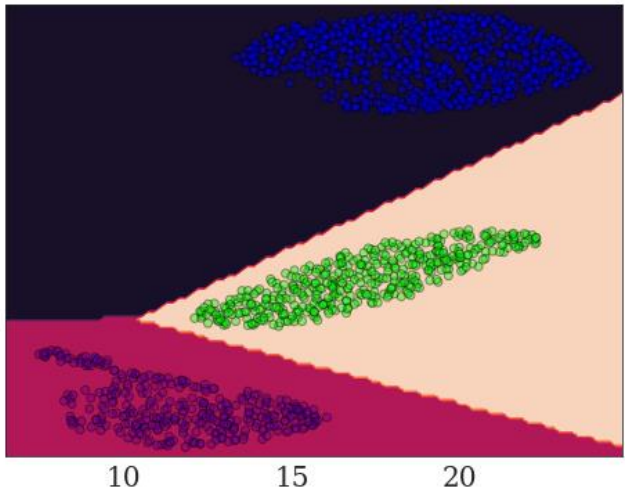
UMAP (8→2 dimensions)

Experiment 1

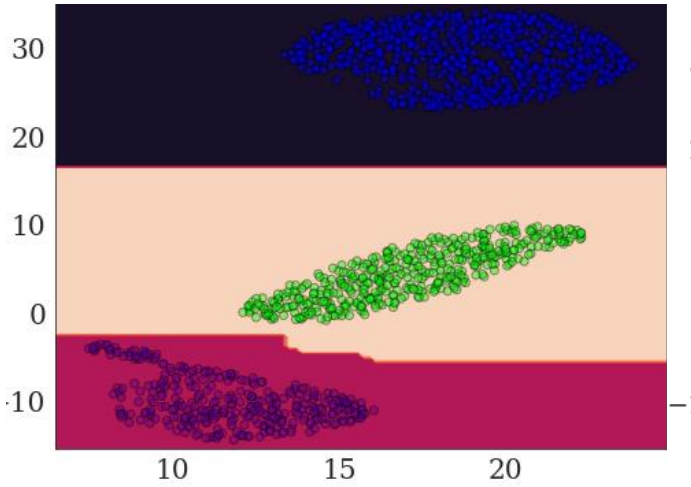
- Reduced from harmonic number 2
- Supervised UMAP



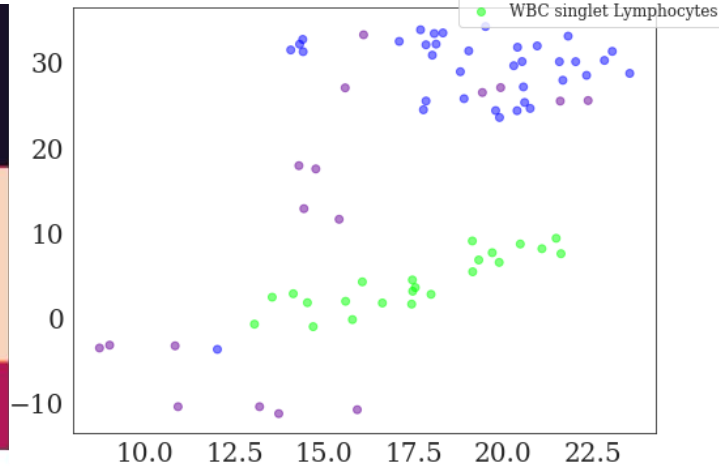
Training set



Linear SVC



Random Forest



Validation set

Comparing classification algorithms

Experiment 1

	Accuracy
Random Forest	0.9
Linear SVC	0.87

- Result for without dimension reduction.
- Random forest method gives slightly better accuracy than Linear SVC.

Comparing dimension reduction methods and how harmonic number affects classifier.

Experiment 1

Using Random Forest classifier:

	Mean of accuracies	Standard deviation
W/o dimensionality reduction	0.915	0.00568
PCA	0.818	0.000620
UMAP	0.867	0.000420

Mean and standard deviation of accuracy values over 2 to 15 harmonic number.

- Small std dev → large harmonic numbers ≠ better classification
- No dimension reduction yields best classification outcome followed by UMAP then PCA

Experiment 1 Summary

Results

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF LIGHT

FAU

- FDs is a good feature to differentiate sRBCs from Lymphocytes.
- Classification of singlet RBC focused and unfocused not perfect.
- Chosen harmonic number N=2 is more than sufficient for classification.
- No dimension reduction gives best classification accuracy.
- Random Forest offers only slightly better accuracy than Linear SVC.

Recap

Defining the boundaries of the thesis.

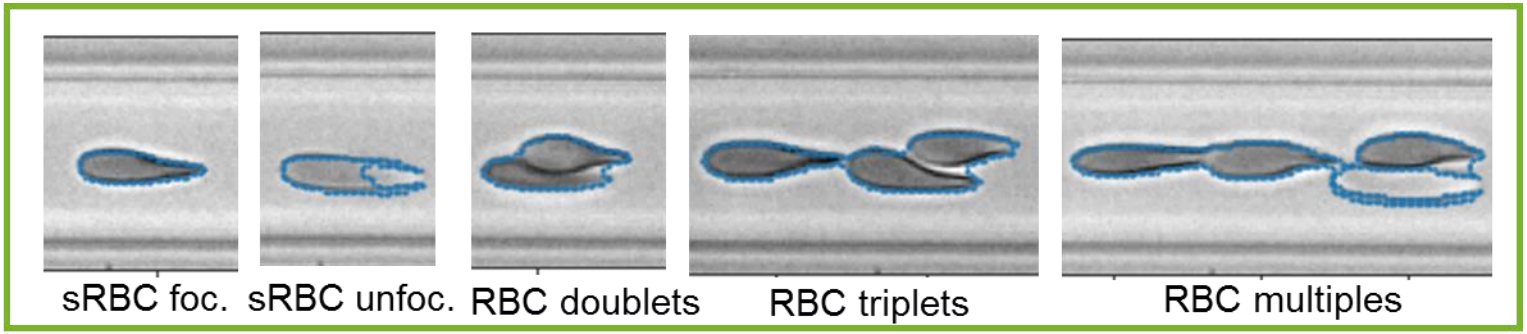
MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

experiments

1

2



3

Classify between control versus long covid.

Tested 15 blood cell types.

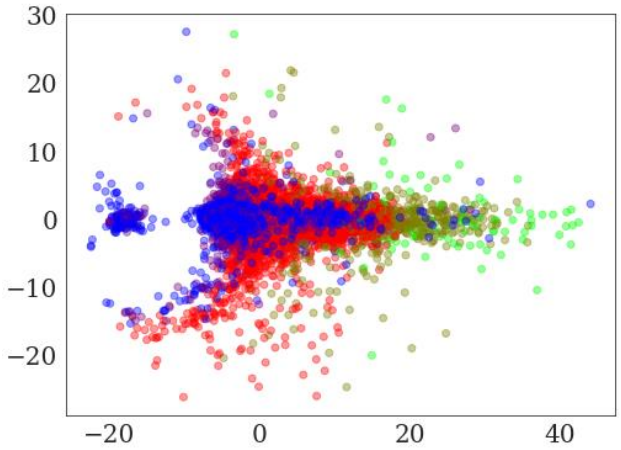
PCA (8→2 dimensions)

Experiment 2

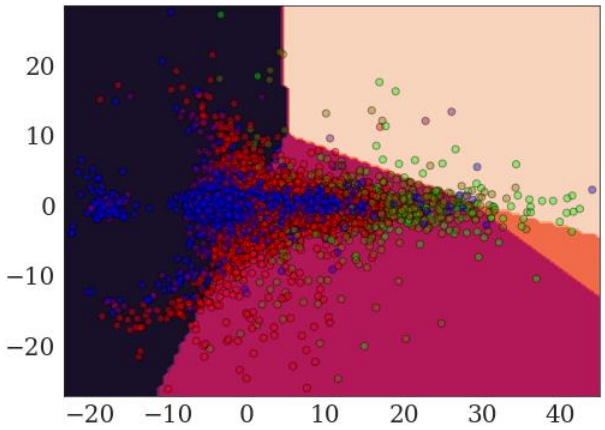
MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

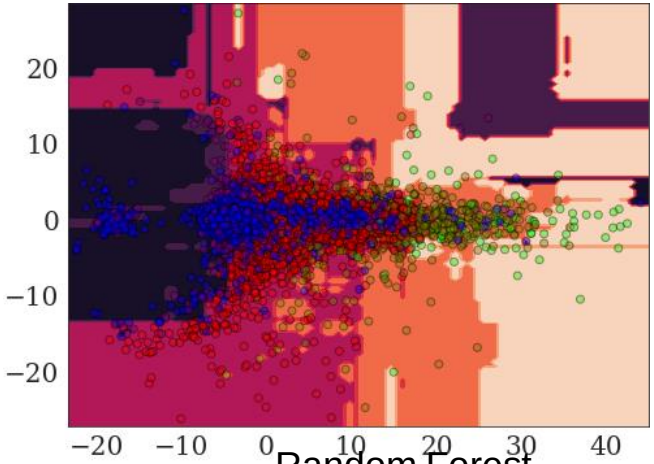
- Reduced from harmonic number 2
- PCA



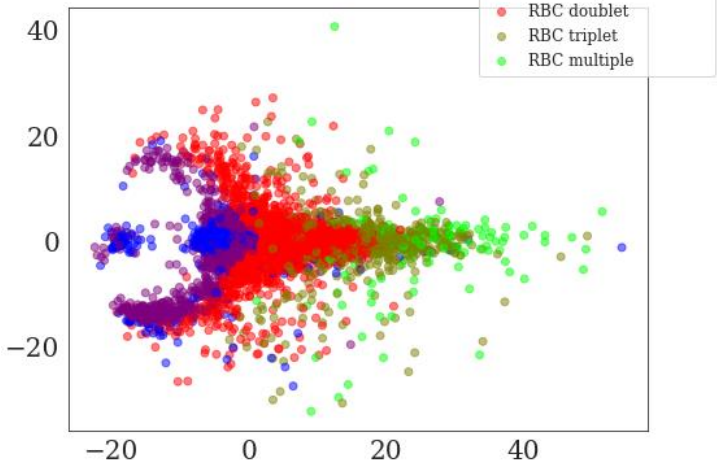
Training set



Linear SVC
(Accuracy = 0.73)



Random Forest
(Accuracy = 0.76)



Validation set

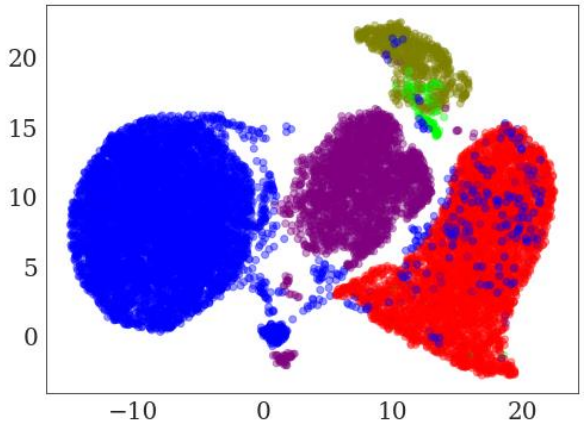
UMAP (8→2 dimensions)

Experiment 2

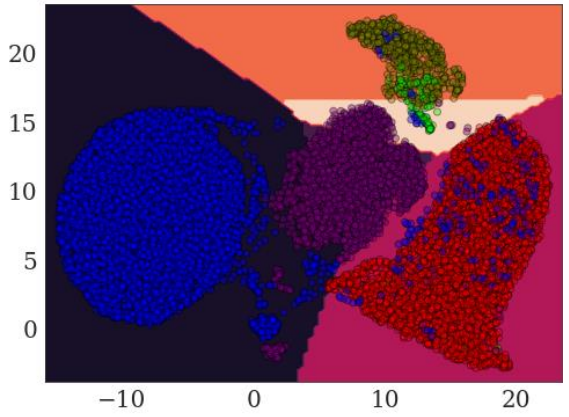
MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

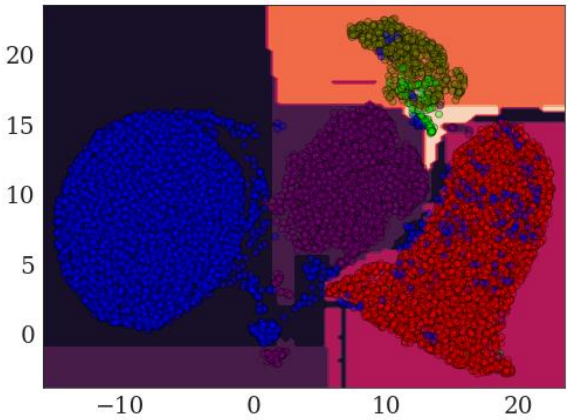
- Reduced from harmonic number 2
- Supervised UMAP



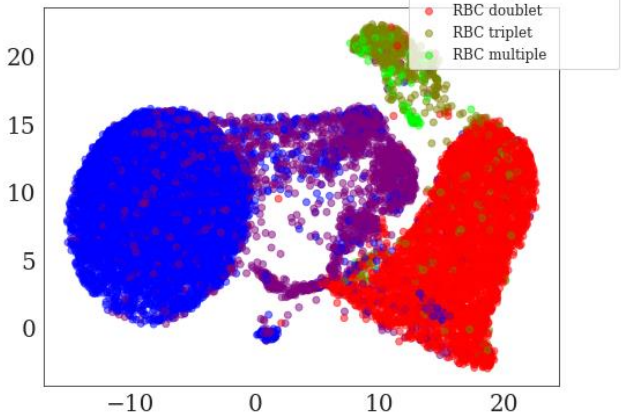
Training set



Linear SVC
(Accuracy = 0.81)



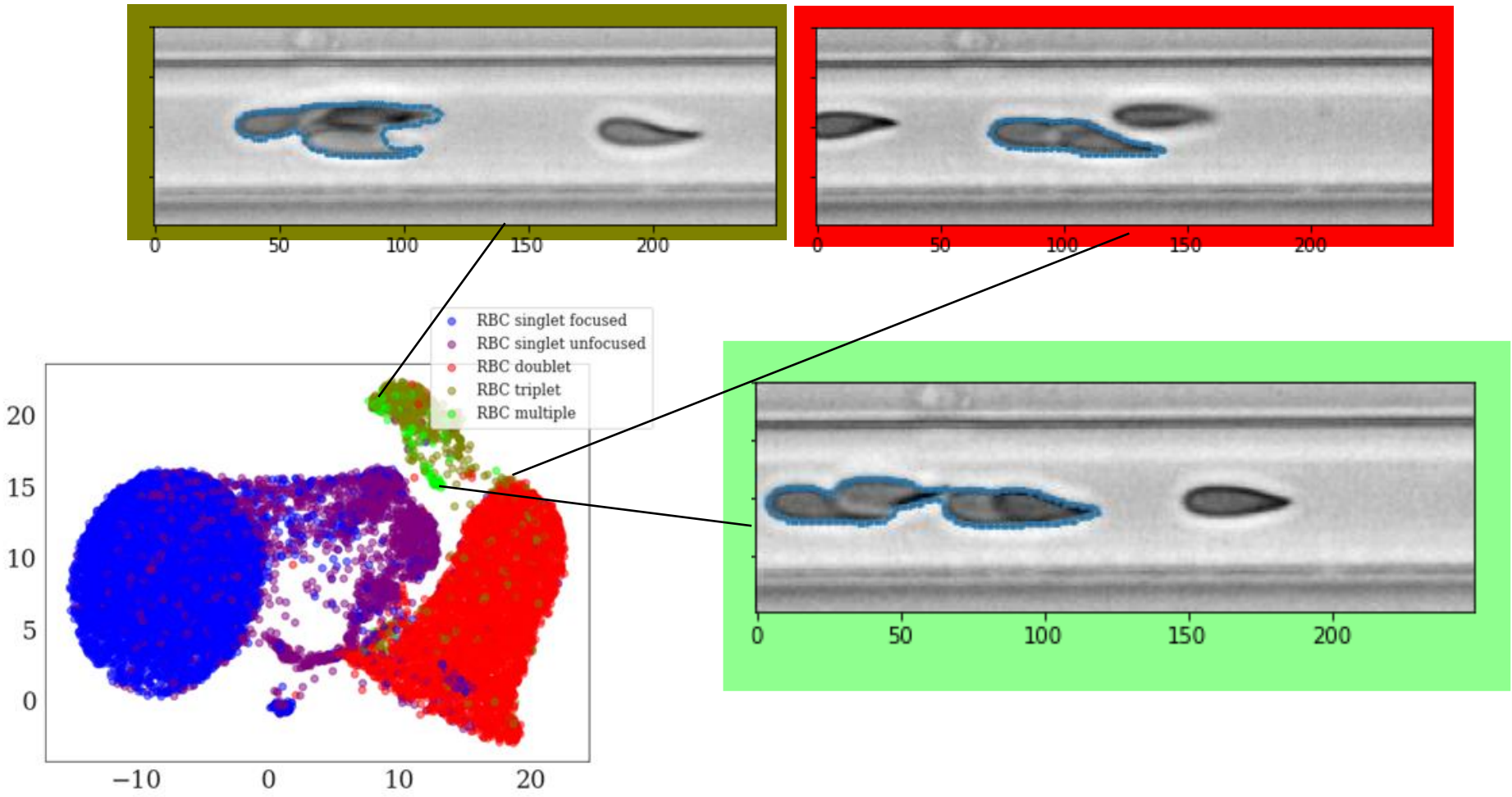
Random Forest
(Accuracy = 0.83)



Validation set

RBC multiples can sometimes have contours that resemble triplets.

Experiment 2



Comparing dimension reduction methods and how harmonic number affects classifier.

Experiment 2

MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

Using Random Forest classifier:

	Mean of accuracies	Standard deviation
W/o dimensionality reduction	0.843	0.00407
PCA	0.764	0.000268
UMAP	0.835	0.000768

Mean and standard deviation of accuracy values from harmonic 2 to 27.

- Small std → large harmonic numbers ≠ better classification
- w/o dimension >UMAP > PCA
- UMAP has similar accuracy to w/o dimension reduction

How classifier method affects classification and reduced space affects performance.

MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

Experiment 2

Random Forest

Accuracy Dimension reduction method	Embedded space 2d	3d	4d	5d	8d
PCA	0.76	0.79	0.82	0.83	-
UMAP	0.83	0.83	0.83	0.83	-
w/o	-	-	-	-	0.84

Linear SVC

Accuracy Dimension reduction method	Embedded space 2d	3d	4d	5d	8d
PCA	0.73	0.73	0.78	0.78	-
UMAP	0.83	0.81	0.82	0.82	-
w/o	-	-	-	-	0.79

Experiment 2 Summary

Results

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF LIGHT

FAU

- FDs are good features for RBC multiplets classification.
- No dimension reduction with Random Forest again yields the best classification.
- UMAP yields similar classification outcome to no dimension reduction already in 2d, and is good for visualization.
- RBC multiples are often confused with triplets regardless of method.

Recap

Defining the boundaries of the thesis.

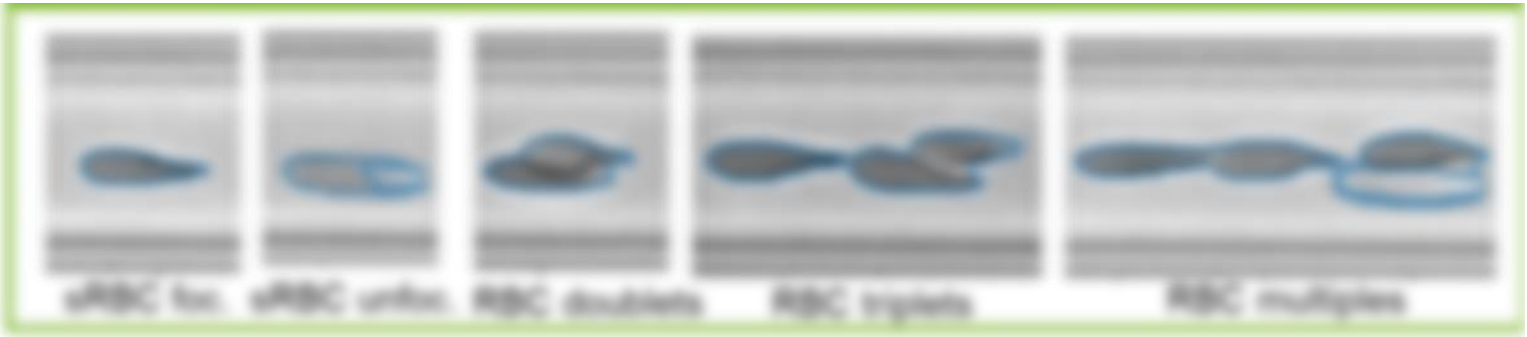
MAX PLANCK INSTITUTE FOR THE SCIENCE OF LIGHT

FAU

experiments

1

2



3

Classify between control versus long covid.

Tested 15 blood cell types.

No learning was observed in Random Forest

Experiment 3

	Random Forest		
	no dim	PCA	UMAP
RBC singlet focused RBC singlet unfocused RBC doublet RBC triplet RBC multiple WBC singlet Lymphocytes Thrombocyte singlet focused Thrombocyte singlet unfocused WBC singlet Neutrophil WBC singlet Eosinophil WBC singlet Basophil WBC singlet Monocytes Thrombocyte multiple Mixed cell doublet Mixed cell triplet	Balanced Accuracy mean<0.5		

- Mean of balanced accuracy and standard deviation from harmonic number 2 to 29.
- Balanced Accuracy Mean < 0.5
- Randomly assigning long –covid and control groups can yield better results.

No learning was observed in Linear SVC

Experiment 3

	Linear SVC		
	no dim	PCA	UMAP
RBC singlet focused	Balanced Accuracy mean<0.5		
RBC singlet unfocused			
RBC doublet			
RBC triplet			
RBC multiple			
WBC singlet Lymphocytes			
Thrombocyte singlet focused	0.55 <Balanced Accuracy mean<0.6		
Thrombocyte singlet unfocused			
WBC singlet Neutrophil			
WBC singlet Eosinophil			
WBC singlet Basophil			
WBC singlet Monocytes			
Thrombocyte multiple			
Mixed cell doublet			
Mixed cell triplet			

- Mean of balanced accuracy and standard deviation from harmonic number 2 to 29.
- Balanced Accuracy Mean < 0.5
- Randomly assigning long –covid and control groups can yield better results.
- Outliers? (in red)

Experiment 3 Summary

Results

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF LIGHT

FAU

- Fourier Descriptors are not good features to classify control versus long covid groups.
- No learning took place.
- Control cells and long covid cells look alike.

Conclusion

Conclusion

- FDs are only suitable if shapes of cell are highly different from each other.
- Worked rather well for RBC multiplets, promising extension to another classification pipeline.
- Overall no dimension reduction gave the best classification outcome, although UMAP also performs sufficiently well and can allow us to visualise the plots.
- Random Forest method is robust as a classification algorithm choice.
- Large number of harmonics does not improve classification.

Vielen Dank
für Ihre Aufmerksamkeit!

Acknowledgements

Professors:
Prof. Jochen Guck
Prof. Vasily Zaburdaev

Main Advisors:
Eoghan O' Connell
Maximilian Schloegel
Paul Müller
Wolfram Pönisch

RTDC Advisors:
Shada Abuhattum Hofemeier
Felix Reichel
Marketa Kubankova
Martin Kräter

Admin organisation:
Yeo Gee Hye
Silke Besold

Providing tips to comply with the masters degree curriculum:
Miriam Schnitzerlein
Tim Klingberg

All members of the Guck Lab for leaving me with wonderful memories while at the Schloss this year.