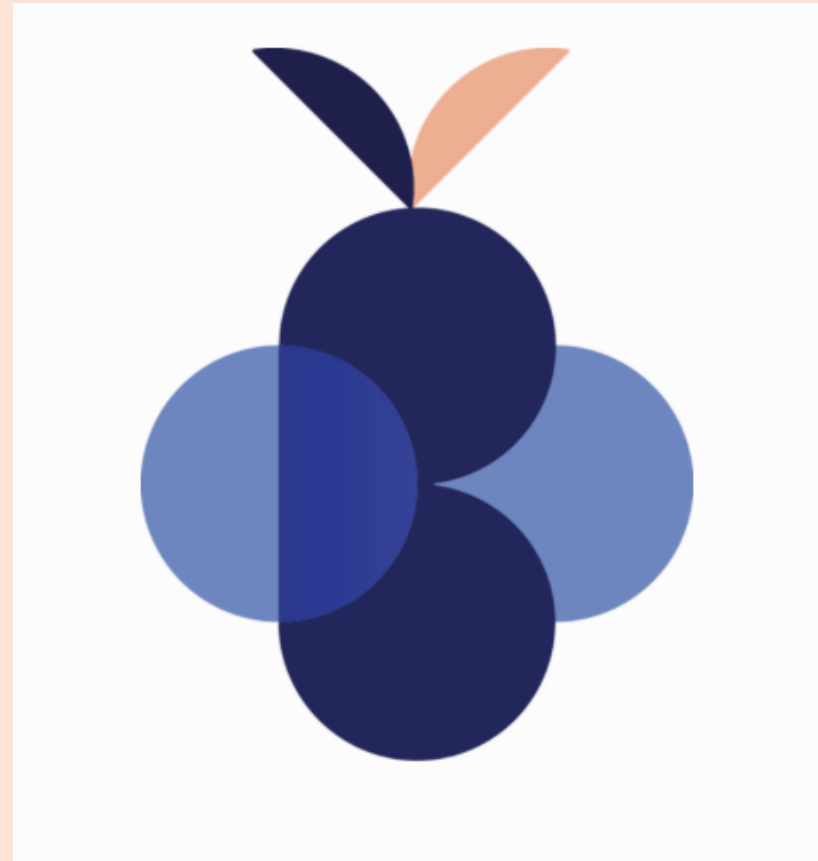


# Bumblekite

Machine learning summer school in  
health, care and biosciences

ETH Zürich, Switzerland



# General information

- Annual machine learning summer school
- Target audience:
  - Early career professionals
  - Students (graduate)
- A dynamic schedule:
  - Lectures
  - Tutorials
  - Leadership conversation series
  - Communication sessions



## bumblekite machine learning summer school 2024

(final schedule, updated: 2nd July 2024)

June 30th, Sun	July 1st, Mon	July 2nd, Tue	July 3rd, Wed	July 4th, Thu	July 5th, Fri	July 6th, Sat
		08:45 Intro to the day	08:45 Intro to the day		08:45 Intro to the day	
	09:30 Registration 09:45 Welcome session	09:00 Siley Ba	09:00 Valeria De Luca	08:30 Kickboxing with Leo 08:30 Group breakfast with Drago	09:00 Ece Özkan Elsen	09:00 - 15:00 Datathon at University Hospital Balgrist
	10:30 - 11:00 Break	10:30 - 11:00 Break	10:30 - 11:00 Break	10:30 - 11:00 Break	10:30 - 11:00 Break	
	11:00 ETH tour	11:00 Paul Clemencon	11:00 Paul Clemencon	11:00 Lito Kriara	11:00 Krishna Chaitanya, Pushpak Pati	11:30 - 14:00 Swimming and picnic at Lake Zürich
	12:00 - 13:00 Lunch	12:00 Antonija Burcul, Ashwarya Parthasarathy, Nina Sesto	12:00 Drago Plečko		12:15 - 13:15 Lunch	
	13:00 Leo Celi	13:00 - 14:00 Lunch	13:00 - 14:00 Lunch	13:00 - 14:00 Lunch	13:15 Krishna Chaitanya, Pushpak Pati	
	14:30 - 15:00 Break	14:00 Leo Celi, Aya El Mir	14:00 Drago Plečko	14:00 Lito Kriara	15:15 - 15:45 Break	
15:30 - 18:00 Hike to Uetliberg	15:00 Closing the day	16:15 - 16:45 Break		16:00 - 16:30 Break	15:45 - 16:45 Antonija Burcul, Tracy Glass, Rebecca Kaufmann	
	15:30 - 17:00 Janet Adeyemi, Leo Celi, Anja Hartewig, Tobias Heimann, Kortine Kleinheinz	16:45 - 17:45 Luca Finelli	17:00 - 17:30 Break	18:30 Closing the day	16:45-17:30 Farewell	
18:00 - 19:00 Guided city tour	18:00 - 20:00 Social evening	17:45 - 19:30 Leo Celi, Aya El Mir	17:30 Closing the day	17:00 - 18:00 Maria Cervera de la Rosa, Shez Partovi, Fabian Rudolf	18:00 Group dinner with Jennifer	
		19:30 Closing the day	18:00 - 19:00 Gian-Reto Grend, Haris Shuaib, Bettina Wäpf			
		20:15 Group dinner with Ece	20:00 Group dinner with Valeria	19:00 Group dinner with Jonas		

# Lectures

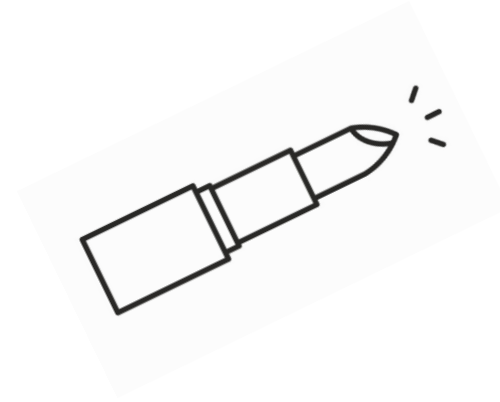
## Bias in AI

- Data is context dependent
- The need to be critical in ML
- Expert knowledge
- Interdisciplinarity



## ML in L'OREAL

- Skin tone prediction (CNNs)
- Foundation & lipstick recommendation (CAGAN)
- Skin condition prediction
- Problems with sensor data



# Lectures (2)

## ML in Novartis

- Use of clinical omics
- Disease markers
- Parkinson's disease, kidney disease, etc.

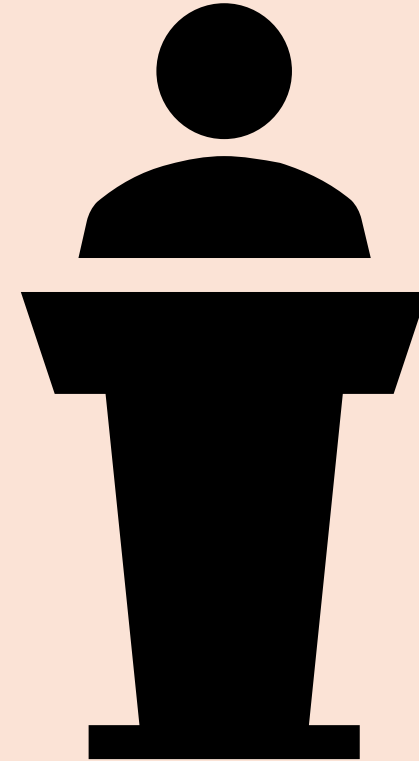


## ML in Roche

- Prioritizing fairness & interpretability
- Debiasing deep chest x-ray classifiers
- Prediction of pulmonary hypotension in newborns
- Prediction of appendicitis

# Communiation lecture

- Communication & business development
- **Importance of negotiation**
  - Value vs price
  - Buy when value > price
- Case study

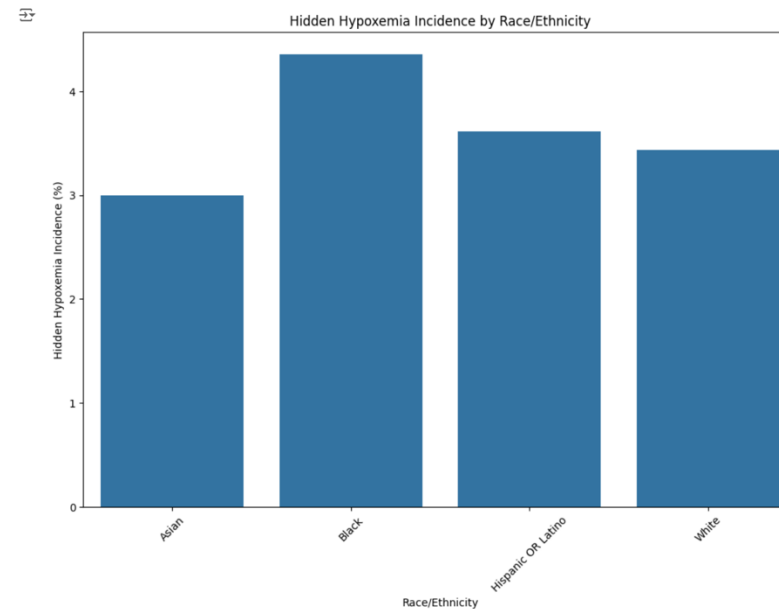
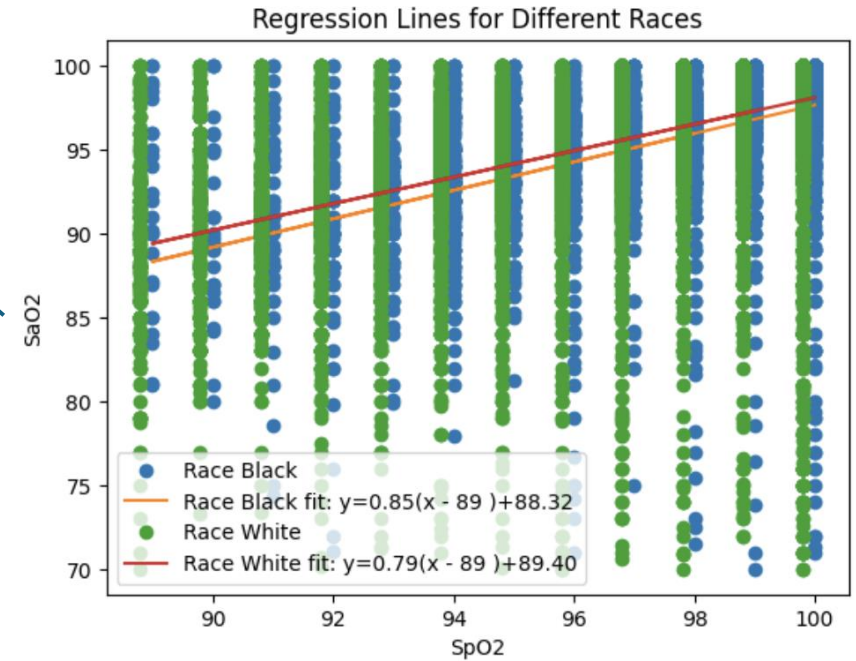


# Tutorials

## Bias in AI

- BOLD dataset
- Bias in pulse oximetry accuracy
  - Biased towards darker-skinned patients
  - Hidden hypoxemia
- Predicting in-hospital mortality
- Emphasis on examining bias, not ML

An invasive measurement of oxygen saturation directly from the blood using an arterial blood gas test.



A non-invasive method used to measure the oxygen level in the blood. It is usually done using a device clipped onto a finger, toe, or earlobe.



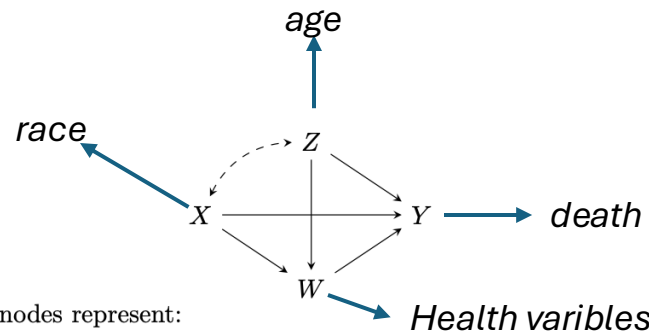
	features	race_ethnicity	auc	Weighted Precision	Weighted Recall	Weighted F1-Score	balanced_accuracy
3	Exaggerated SpO2	Asian	0.727519	0.852817	0.860759	0.819215	0.576241
3	SaO2	Asian	0.726617	0.840190	0.854430	0.807050	0.556241
3	SpO2	Asian	0.728421	0.852817	0.860759	0.819215	0.576241
1	Exaggerated SpO2	Black	0.758489	0.797087	0.832432	0.794723	0.572467
1	SaO2	Black	0.759869	0.797797	0.832432	0.796895	0.577410
1	SpO2	Black	0.760139	0.805853	0.836757	0.801089	0.582496
2	Exaggerated SpO2	Hispanic OR Latino	0.782582	0.801163	0.822014	0.783712	0.597269
2	SaO2	Hispanic OR Latino	0.775523	0.792869	0.817330	0.775786	0.585641
2	SpO2	Hispanic OR Latino	0.782923	0.801163	0.822014	0.783712	0.597269
0	Exaggerated SpO2	White	0.789874	0.825862	0.844906	0.808421	0.594704
0	SaO2	White	0.786374	0.825589	0.844640	0.807706	0.593354
0	SpO2	White	0.789897	0.826829	0.845172	0.808235	0.593974

- **Logistic regression** results for each race; different features

# Tutorials (2)

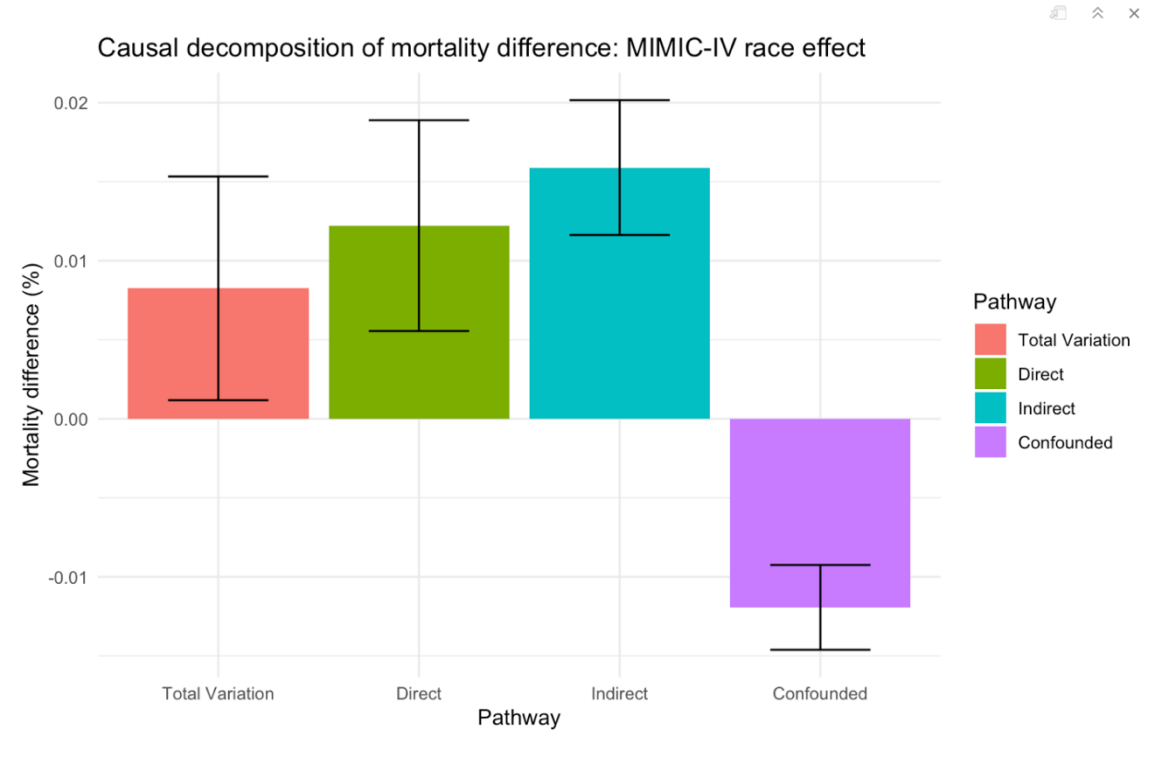
## Causal fairness analysis

- MIMIC-IV dataset
  - Health variables + mortality
- Standard fairness model



where the nodes represent:

- the *protected attribute*, labeled *X* (e.g., gender, race, religion),
- the set of *confounding* variables *Z*, which are not causally influenced by the attribute *X* (e.g., demographic information, zip code),
- the set of *mediator* variables *W* that are possibly causally influenced by the attribute (e.g., educational level or other job-related information),
- the *outcome* variable *Y* (e.g., admissions, hiring, salary).



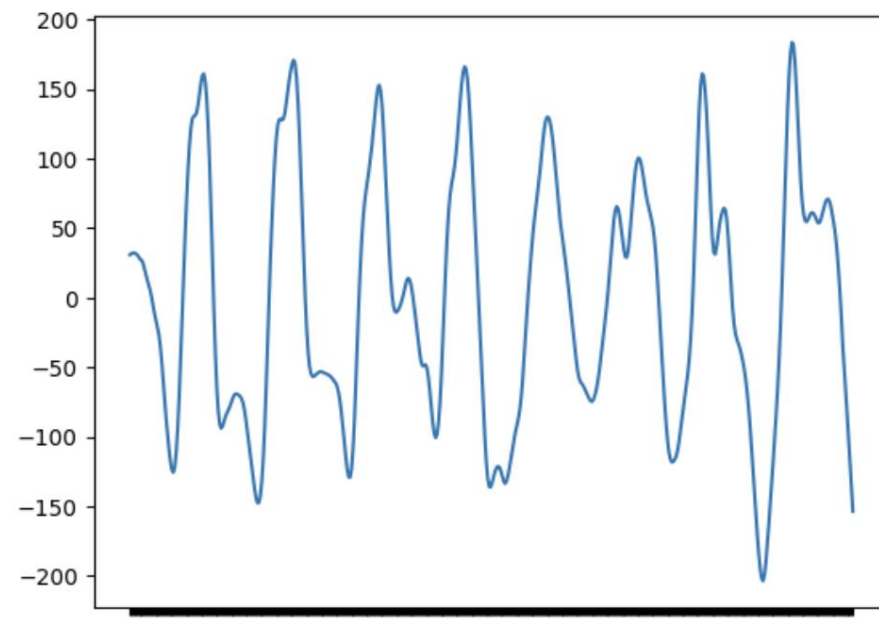
[Want to learn more?](#)



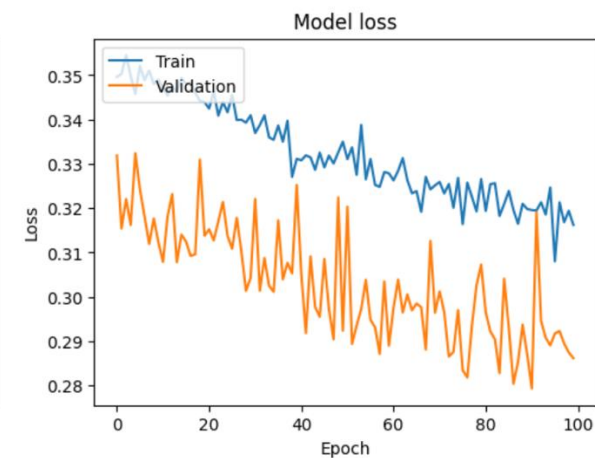
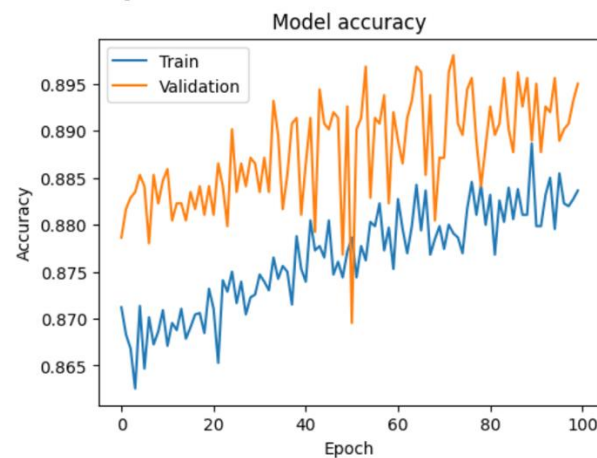
# Tutorials (3)

## Noise in data

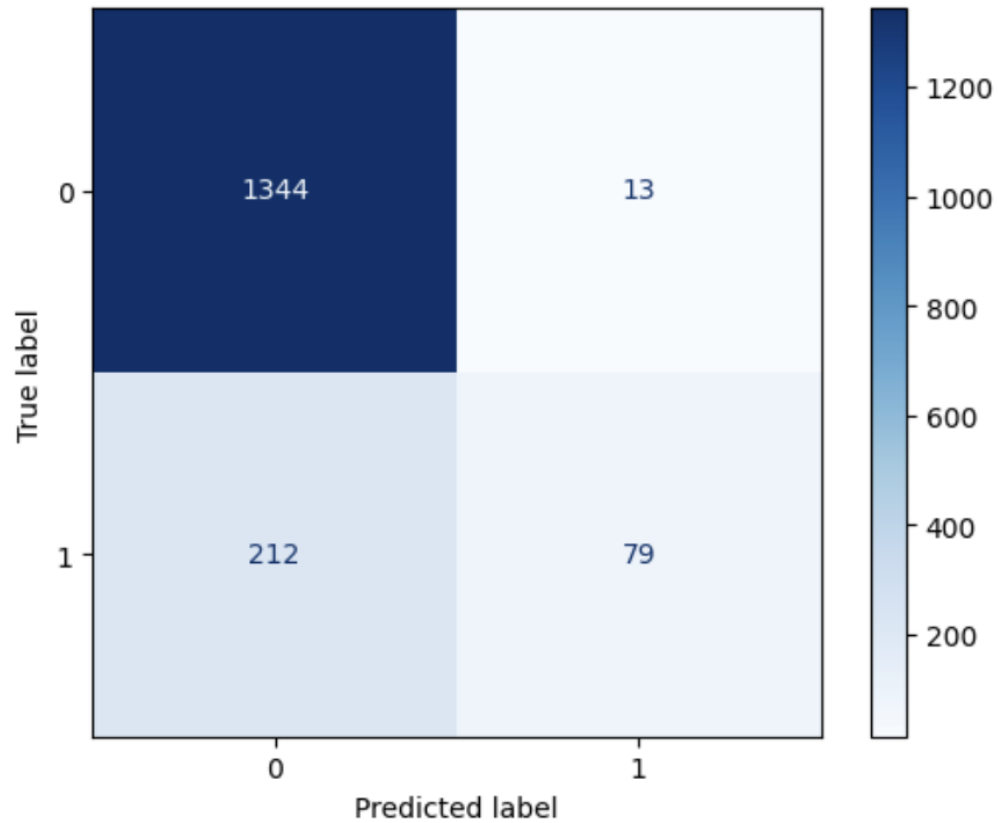
- Deep learning (CNN)
- PPG data
- Dependent instances
- Predicting noisiness of the data
- High accuracy, low/decreasing loss



```
206/206 [=====] - 6s 27ms/step - loss: 0.3194 - accuracy: 0.8827 - val_loss: 0.2875 - val_accuracy: (
Epoch 100/100
206/206 [=====] - 7s 36ms/step - loss: 0.3162 - accuracy: 0.8836 - val_loss: 0.2862 - val_accuracy: (
52/52 [=====] - 1s 10ms/step - loss: 0.3656 - accuracy: 0.8635
Test Loss: 0.36562612652778625
Test Accuracy: 0.8634708523750305
```

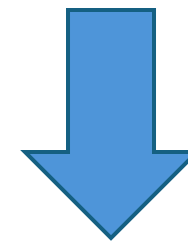


52/52 [=====] - 0s 9ms/step



### After examining the confusion matrix ...

- Model classifies into category 0
- Train data with imbalanced labels



Solutions?

# Social activities, networking

- A lot of events
- Dinners with lecturers
- Community building

