



INSTITUTE



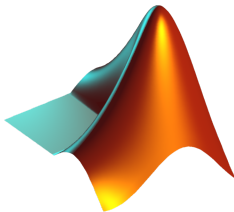
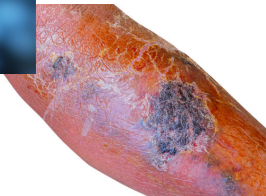
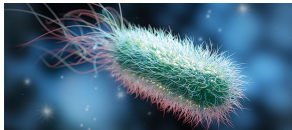
OF SCIENTIFIC INSTRUMENTS

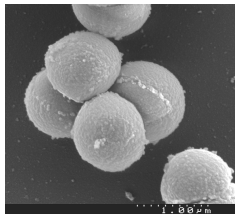
The Czech Academy of Sciences

Jak pomocí světla a MATLABu rozlišit bakterie

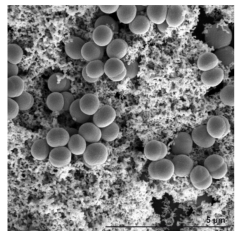
Martin Šiler

Ústav přístrojové techniky AV ČR, v.v.i.
Královopolská 147, 612 64 Brno

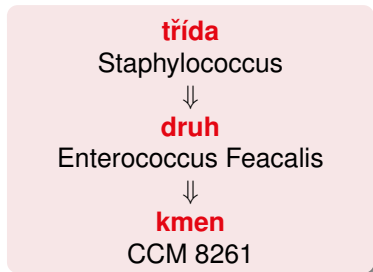




- **Staphylococcus aureus**
- Toxická bakterie
- Krevní infekce
- Odolná na antibiotika

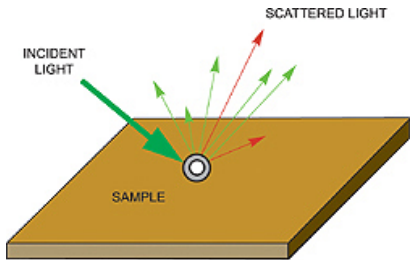


- **Staphylococcus epidermidis**
- běžná na kůži
- převážně neškodná
- Ale tvoří biofilm

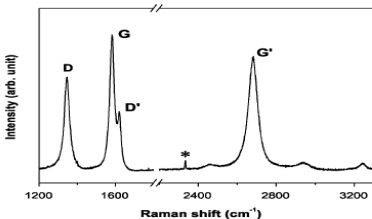
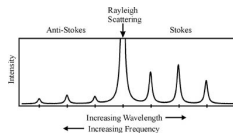
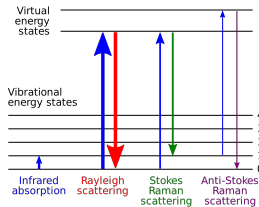


¹ K. Rebrošová et al., Future Microbiology **12**, 10 (2017).

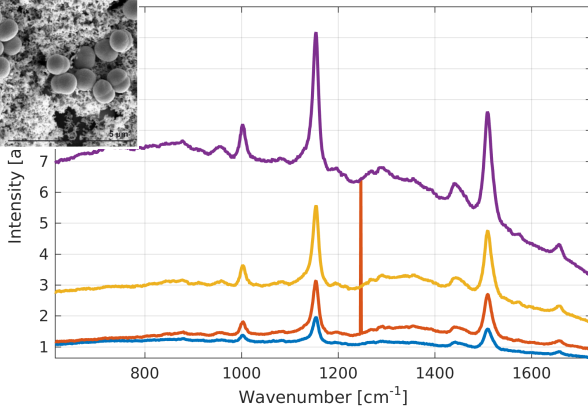
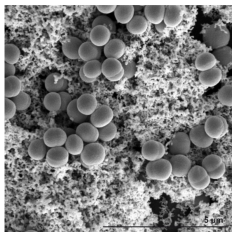
² K. Rebrošová et al., Sci. Rep. **7**, 14846 (2017).



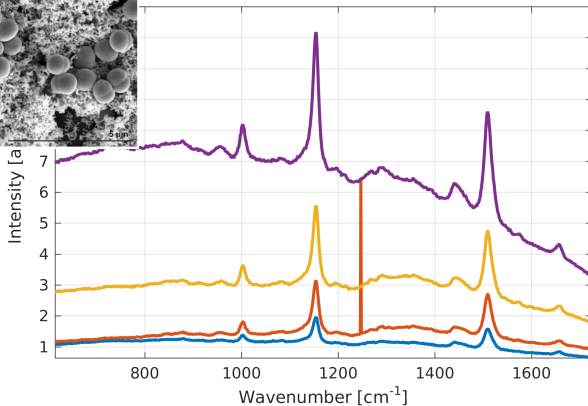
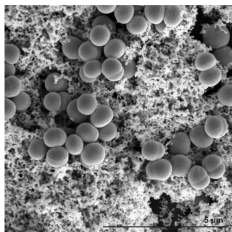
- RAYLEIGH SCATTER (SAME WAVELENGTH AS THE INCIDENT LIGHT)
- RAMAN SCATTER (NEW WAVELENGTH)



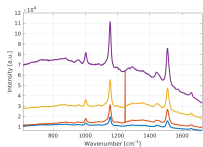
- Nepružný rozptyl světla
- zakódovaná chemická struktura
- slabý



- **Různá škála**
poloha měření v kolonii
- **Poissonův šum**
aditivní, ale $\sigma \sim \sqrt{\mu}$
- **Fluorescence**
jiný rozptyl
silnější
trochu jiný charakter
- **artefakty**
“kosmické paprsky”



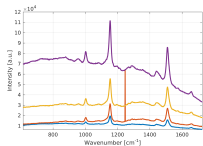
- 28 druhů bakterií
- pro každý 3-60 kmenů (~305 celkem)
- 8-15 spekter pro kmen
- spektrum 1013 bodů (dimenzí)
- nezávislá klasifikace hmotnostní spektroskopii (pomalá)



klasifikace

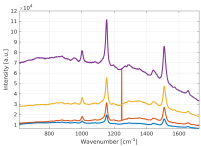
Přímá cesta: (konvoluční) neuronové sítě

- není dost dat
- nevyvážené třídy
- obří variabilita na vstupu

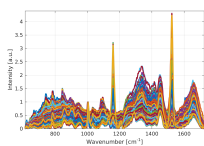


klasifikace

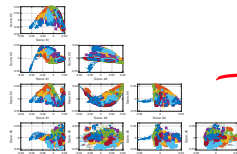
Po krůčcích



filtrování

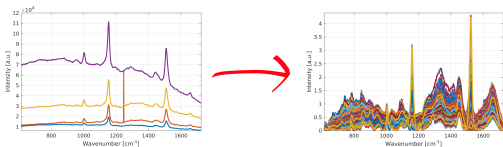


příznaky



klasifikace

	s. aureus	s. bacillans	s. bacillis	s. capitis	s. epidermidis	s. haemolyticus	s. hominis	s. pseudointestinalis	s. saprophyticus	s. salivarius	s. simulans	s. warneri	s. xylinus	Total Predicted	Total True
s. aureus	100													100	100%
s. bacillans		63	1											64	100%
s. bacillis									5		1			6	10%
s. capitis				64										64	100%
s. epidermidis					622	1								623	100%
s. haemolyticus						213								213	100%
s. hominis			1	1	204				2		1			209	95%
s. pseudointestinalis							82							82	100%
s. saprophyticus			1				128							129	100%
s. salivarius								162						162	100%
s. simulans			1	1				148						150	97%
s. warneri									79		1			80	100%
s. xylinus													79	79	100%
Positive Predictive Value	100%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%
False Discovery Rate	0%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%

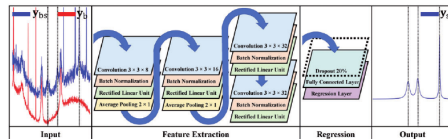


Klasický přístup

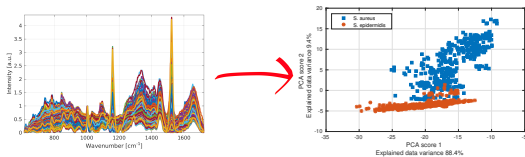
- Savitzky–Golay
- **fluorescenční pozadí**
Valící se kruh¹
Iterativní polynomy²
- normalizace intenzity

Neuronové sítě³

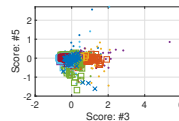
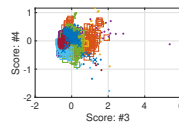
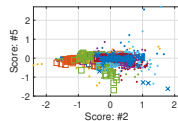
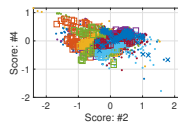
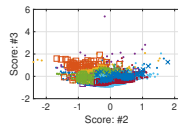
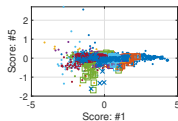
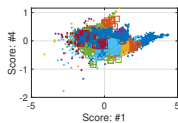
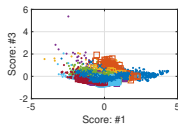
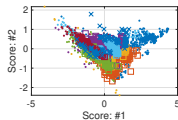
- můžeme generovat umělá data
- pokus loni, neúspěšný
- **Diplomová/bakalářská práce**



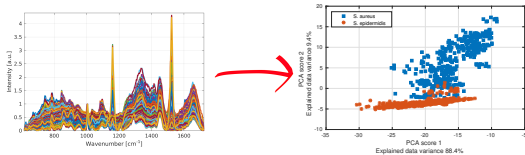
¹ N. N. Brandt et al., Appl. Spectrosc. **60**, 288–293 (2006).
² K. H. Liland et al., Appl. Spectrosc. **64**, 1007–1016 (2010).
³ J. Wahl et al., Applied Spectroscopy **74**, 427 (2020).



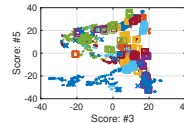
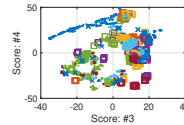
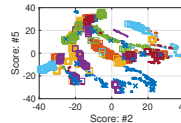
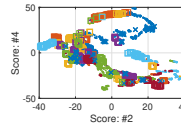
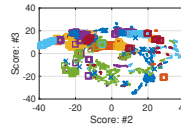
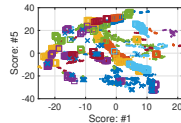
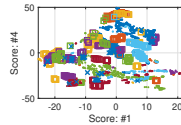
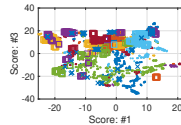
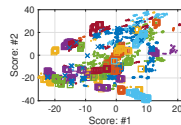
- Nahradit spektrum několika body
- Principal Component Analysis
 - 1 centrovaná × necentrovaná
 - 2 lineární × nelineární
- Non-negative matrix factorization
- tSNE, UMAP
- Kolik příznaků



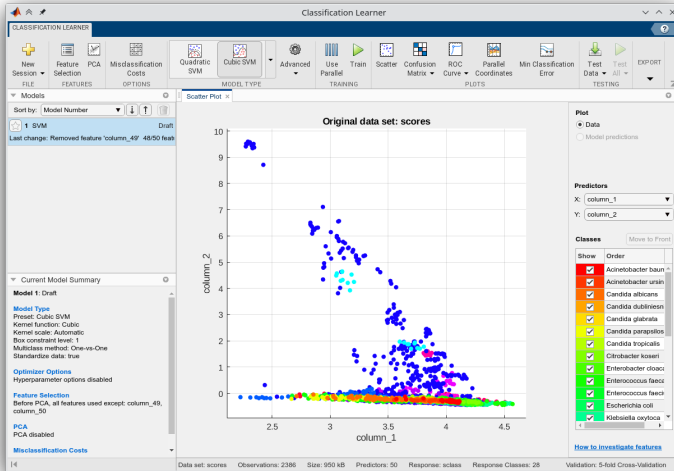
cPCA



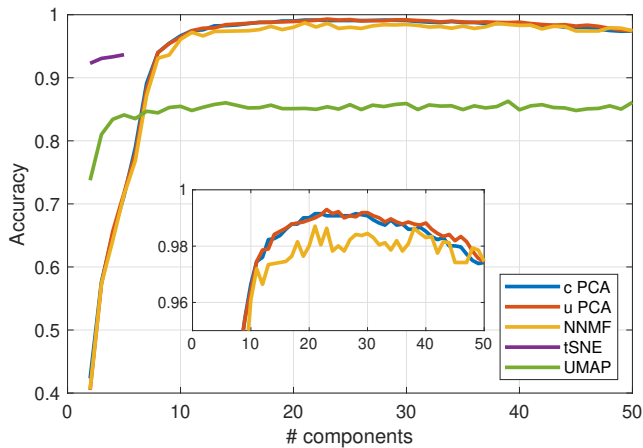
- Nahradit spektrum několika body
- Principal Component Analysis
 - 1 centrovaná × necentrovaná
 - 2 lineární × nelineární
- Non-negative matrix factorization
- tSNE, UMAP
- Kolik příznaků



tSNE



- Metody: kNN, SVM (LDA, Random Forest)
- Cross-validace



třída
⇓
druh
⇓
kmen
⇓
spektrum

- Trénovací × testovací data
- 8-15 spekter na “vzorek”
- Klasifikátor po spektrech



- 1 druh výrazně dominuje
- 1 druh dominuje
- není možné rozhodnout

	<i>Acinetobacter baumannii</i>	<i>Acinetobacter ursingii</i>	<i>Candida albicans</i>	<i>Candida dubliniensis</i>	<i>Candida glabrata</i>	<i>Candida parapsilosis</i>	<i>Candida tropicalis</i>	<i>Citrobacter koseri</i>	<i>Enterobacter cloacae</i>	<i>Enterococcus faecalis</i>	<i>Enterococcus faecium</i>	<i>Escherichia coli</i>	<i>Klebsiella oxytoca</i>	<i>Klebsiella pneumoniae</i>	<i>Micrococcus luteus</i>	<i>Proteus vulgaris</i>	<i>Providencia stuartii</i>	<i>Pseudomonas aeruginosa</i>	<i>Serratia marcescens</i>	<i>Staphylococcus aureus</i>	<i>Staphylococcus epidermidis</i>	<i>Staphylococcus haemolyticus</i>	<i>Staphylococcus hominis</i>	<i>Staphylococcus lugdunensis</i>	<i>Staphylococcus pasteuri</i>	<i>Staphylococcus warneri</i>	<i>Streptococcus agalactiae</i>	<i>Streptococcus oralis</i>	Unidentified strain
<i>Acinetobacter baumannii</i>	1																												
<i>Acinetobacter ursingii</i>	0																												
<i>Candida albicans</i>			2											1															
<i>Candida dubliniensis</i>				1																									
<i>Candida glabrata</i>					1																								
<i>Candida parapsilosis</i>						1																							
<i>Candida tropicalis</i>							1																						
<i>Citrobacter koseri</i>								0																					
<i>Enterobacter cloacae</i>									1																				
<i>Enterococcus faecalis</i>										5																			
<i>Enterococcus faecium</i>											2																		
<i>Escherichia coli</i>												11																	
<i>Klebsiella oxytoca</i>													0																
<i>Klebsiella pneumoniae</i>														2															
<i>Micrococcus luteus</i>															1														
<i>Proteus vulgaris</i>																1													
<i>Providencia stuartii</i>																	1												
<i>Pseudomonas aeruginosa</i>																		3/1											
<i>Serratia marcescens</i>																			1										
<i>Staphylococcus aureus</i>																				14									
<i>Staphylococcus epidermidis</i>																					6								
<i>Staphylococcus haemolyticus</i>																						0	1						
<i>Staphylococcus hominis</i>																							2						
<i>Staphylococcus lugdunensis</i>																								1					
<i>Staphylococcus pasteuri</i>																									1				
<i>Staphylococcus warneri</i>																										1			
<i>Streptococcus agalactiae</i>																											1		
<i>Streptococcus oralis</i>																												1	

True Class

metoda	PCA	zelená správně	žlutá správně	červená	žlutá špatně	zelená špatně
kNN	centrovaná	67.6%	4.2%	16.9%	4.2%	7.0%
	necentrovaná	67.6%	7.0%	12.7%	5.6%	7.0%
SVM	centrovaná	76.1%	9.9%	11.3%	0.0%	2.8%
	necentrovaná	81.7%	7.0%	4.2%	2.8%	4.2%
	<i>kombinace</i>	80.3%	7.0%	8.5%	0.0%	4.2%

- Pomocí světla a AI rozlišíme bakterie
- Pre-processing a výběr příznaků jsou klíčové
- “Semaforová klasifikace”
- SVM vítězí

■ Příkaz `system` volá externí příkazy

- 1 `cp / copy / mv` třídění na trénovací–testovací dataset, hromadné přejmenování
- 2 `latex / pdflatex` MATLAB generuje \LaTeX zprávu a rovnou generuje finální pdf
- 3 `wmctrl -r "Figure 3" -t 6` v Linuxu přesouvá okna mezi plochami

■ Spline (po částech spojitě polynomy) fce `ppval`

Varování: `histc` is not recommended. Use `histcounts` instead.

■ Memory leak v `parfor` ?

- Simulace `dat` → Vyhodnocení: minimalizace `fminsearch` + složitá funkce s persistentními daty
- 1 GB `dat`, ± 3 minuty, 100 tis. krát
- `parfor` 64 jader, mezi iteracemi se neuvolní nějaké paměť
- původ neznámý (zatím)
- workaround zapínat a zapínat `parpool`
 - náhodný generátor se pokaždé inicializuje stejně



Martin Šiler, K. Rebrošová, ...

www.isibrno.cz

The research was supported by European Structural and Investment Funds Project "Holographic endoscopy for in vivo applications"

(No. CZ. 02.1.01/0.0/0.0/15_003/0000476) and MEYS CR (LO1212).