

Univerzita Palackého v Olomouci  
Přírodovědecká fakulta

Katedra fyzikální chemie



**Studie využitelnosti Ramanovy  
spektroskopie v analýze vybraných  
demografických parametrů na základě  
analýzy klinických vzorků**

BAKALÁŘSKÁ PRÁCE

Autor:	Laura Volejníková
Studijní obor:	Aplikovaná chemie
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Václav Ranc, Ph.D.
Rok:	2023



### **Prohlášení**

Prohlašuji, že jsem tuto práci vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou v seznamu použité literatury. Souhlasím s tím, že práce je prezenčně zpřístupněna v knihovně Katedry fyzikální chemie, Přírodovědecké fakulty, Univerzity Palackého v Olomouci.

V Olomouc dne .....

.....  
Laura Volejníková

## **Poděkování**

Ráda bych poděkovala vedoucímu bakalářské práce doc. RNDr. Václavu Rancovi, PhD. za umožnění zpracování bakalářské práce pod jeho odborným vedením, za jeho skvělý přístup a trpělivost. Velké díky patří mé rodině, která mi umožnila studium na vysoké škole a vždy byla mojí velkou oporou. V neposlední řadě bych ráda poděkovala svým kamarádům za jejich podporu a pozitivní energii, kterou mi pravidelně předávali. Mé největší poděkování patří mému příteli, který tu pro mě byl v horších i lepších chvílích mého studia, ale nikdy se k mým problémům neotočil zády. Tímto mu děkuji za jeho neuvěřitelně velkou podporu, trpělivost a energii.

Laura Volejníková

*Název práce:*

**Studie využitelnosti Ramanovy spektroskopie v analýze vybraných demografických parametrů na základě analýzy klinických vzorků**

*Autor:* Laura Volejníková

*Studijní program:* Aplikovaná chemie

*Druh práce:* Bakalářská práce

*Vedoucí práce:* doc. RNDr. Václav Ranc, Ph.D.

Ústav molekulární a translační medicíny

*Abstrakt:* Ramanova spektroskopie je nedestruktivní neinvazivní analytická metoda, která je univerzální technikou v rámci použití různých typů vzorků v různých oblastech výzkumu. Největší výhodou metody je rychlost měření bez nutnosti úprav vzorků a také možnost přenosného spektrometru, který by mohl být využit v terénu, což nahrává tématu bakalářské práce. Analýzou klinických vzorků s možností využití Ramanovy spektroskopie pro rozdělení souboru vzorků podle pohlaví byly prozatím publikované dvě práce, ale ani jedna s použitím krevní plazmy. Cílem práce bylo vyvinout metodu, která by přispěla do oblasti výzkumu v rámci zpracování spektrálních dat na základě vybraného demografického parametru. Po předúpravách dat, které byly pro práci klíčové z důvodu velkého počtu vzorků (téměř 700) byly za použití PCA, HCA analýz a algoritmu PLS dosaženy zajímavé výsledky. Spektra mužů a žen se povedlo na základě uvedených analýz rozdělit téměř s 95% pravděpodobností.

*Klíčová slova:* Ramanova mikroskopie, klinické vzorky, krevní sérum, plazma

*Title:*

**Study of the application of Raman spectroscopy in the analysis of selected demographic parameters based on the analysis of clinical samples**

*Author:* Laura Volejníková

*Abstract:* Raman spectroscopy is a non-destructive, non-invasive analytical method that is a universal technique in the use of different sample types in various research fields. The greatest advantage of the technique is the speed of measuring without the need for sample modifications and also the possibility of a handheld spectrometer that could be used in the field, which is relevant to the topic of the bachelor thesis. Analyzing clinical samples with the possibility of using Raman spectroscopy to recognize the gender distribution of the sample set, two studies have been published so far, but neither using blood plasma. The aim of this thesis was to develop an algorithm that would contribute to the research field in the processing of spectral data based on a selected demographic parameter. After data pre-processing, which was essential for the work, due to the large number of samples (almost 700), interesting results were obtained using PCA, HCA analysis and the PLS algorithm. Based on these analysis, the spectra of men and women could be separated with almost 95% probability.

*Key words:* Raman microscopy, clinical samples, blood serum, plasma

# Obsah

## Seznamy a značení

Úvod	11
<b>Teoretická část</b>	<b>13</b>
<b>1 Ramanova spektroskopie</b>	<b>13</b>
1.1 Historie metody	13
1.2 Základní teorie	13
1.3 Instrumentace Ramanovy spektroskopie	15
1.4 Ramanovo spektrum	16
<b>2 Ramanova mikroskopie</b>	<b>17</b>
2.1 Ramanova mikroskopie	17
2.2 Instrumentace metody	17
2.3 Obecné využití	18
<b>3 Využití v praxi</b>	<b>19</b>
3.1 Potravinářství	19
3.2 Historie	19
3.3 Forenzní analýza	20
3.4 Lékařská diagnostika	20
<b>4 Metody analýzy klinických vzorků pomocí RS</b>	<b>21</b>
4.1 Materiál, množství vzorků	21
4.2 Proměření	22
4.3 Zpracování	22
4.4 Studie zaměřené na rozpoznání pohlaví	23
<b>Experimentální část</b>	<b>25</b>
<b>Cíl práce</b>	<b>25</b>
<b>5 Vlastní měření spekter</b>	<b>27</b>
5.1 Studium přístupu měření	27
5.2 Měření spekter reálných vzorků plazmy	30
<b>6 Spektrální analýza naměřených dat</b>	<b>33</b>
6.1 Předúprava naměřených dat k využití v dalších analýzách	33
6.1.1 Vybrání spektrální oblasti	34
6.1.2 Odstranění spektrálního pozadí	34

---

6.1.3	Normalizace dat . . . . .	35
6.1.4	Odstranění šumu . . . . .	35
6.2	Další zpracování naměřených dat . . . . .	36
6.2.1	Analýza hlavních komponent – PCA . . . . .	36
6.2.2	Analýza hierarchického shlukování – HCA . . . . .	36
6.2.3	Algoritmus nejmenších parciálních čtverců – PLS . . . . .	36
6.2.4	ROC grafy . . . . .	36
<b>7</b>	<b>Zhodnocení naměřených dat</b>	<b>39</b>
7.1	Vybrání vhodné metody pro odstranění spektrálního pozadí . . . . .	39
7.1.1	Lowpass . . . . .	39
7.1.2	Modpolyfit . . . . .	40
7.2	Volba využití funkce pro normalizaci dat . . . . .	41
7.2.1	Využití funkce PQN a TotInt s použitím metody lowpass . . . . .	41
7.2.2	Využití funkce PQN a TotInt s použitím metody modpolyfit . . . . .	42
7.3	Zvolení vhodného derivačního stupně . . . . .	43
7.4	Další zpracování dat po jejich předešlé úpravě . . . . .	46
7.4.1	Analýza hlavních komponent – PCA . . . . .	46
7.4.2	Clustering . . . . .	47
<b>8</b>	<b>Diskuze a zhodnocení dosažených výsledků</b>	<b>49</b>
	<b>Závěr</b>	<b>51</b>
	<b>Summary</b>	<b>53</b>
	<b>Bibliografie</b>	<b>55</b>





# Seznamy a značení

## Seznam obrázků

1.1	Typy vibračních molekul . . . . .	14
1.2	Energie rozptýlených fotonů, upraveno z [12] . . . . .	15
1.3	Schéma Ramanovy spektroskopie . . . . .	15
2.1	Instrumentace Ramanovy mikroskopie . . . . .	18
4.1	Grafické znázornění podstaty bakalářské práce . . . . .	25
5.1	Ramanská mapa průměrné Ramanské intenzity . . . . .	28
5.2	Univariální analýza Ramanské mapy . . . . .	29
5.3	Průměrná spektra na Ramanské mapě . . . . .	29
5.4	Zastoupení mužů a žen v analýze . . . . .	30
5.5	Kapka plazmy . . . . .	31
6.1	Naměřené spektrum plazmy s označením MUZ_3433_37 . . . . .	34
7.1	Zhodnocení naměřených dat pomocí metody <code>lowpass</code> . . . . .	40
7.2	Zhodnocení naměřených dat pomocí metody <code>modpolyfit</code> . . . . .	41
7.3	Srovnání ROC grafu a AUC hodnot při použití funkce <code>lowpass</code> . . . . .	42
7.4	Srovnání ROC grafu a AUC hodnot při použití funkce <code>modpolyfit</code> . . . . .	42
7.5	Srovnání vyhlazovací metody při použití klouzavého průměru a při použití první a druhé derivace . . . . .	43
7.6	Srovnání ROC grafů a průměrných spekter s využitím první derivace při metodách <code>lowpass</code> a <code>modpolyfit</code> . . . . .	44
7.7	Srovnání zvoleného počtu komponent s použitím funkce <code>lowpass</code> . . . . .	46
7.8	Srovnání zvoleného počtu komponent při použití funkce <code>modpolyfit</code> . . . . .	46
7.9	PCA analýza s použitím metody <code>modpolyfit</code> . . . . .	47
7.10	Srovnání klastrování při použití <code>lowpass</code> metody s využitím klouzavého průměru a <code>modpolyfit</code> s vyhlazením první derivací . . . . .	48

## Seznam tabulek

5.1	Technické parametry měření . . . . .	31
7.1	Souhrnná tabulka s AUC hodnotami po předúpravě dat při použití rozdílných metod pro odstranění spektrálního pozadí a při použití funkcí při normalizaci dat . . . . .	45

## Seznam zkratek

---

Zkratka	Význam
<b>ANN</b>	Umělé neuronové sítě (Artificial Neural Network)
<b>AUC</b>	Plocha pod křivkou (Area Under the Curve)
<b>HCA</b>	Analýza hierarchického shlukování (Hierarchical Clustering Analysis)
<b>IR</b>	Infračervená spektroskopie (Infrared Spectroscopy)
<b>IUPAC</b>	Mezinárodní unie pro čistou a užitou chemii (International Union of Pure and Applied Chemistry)
<b>NMR</b>	Nukleární magnetická rezonance
<b>PCA</b>	Analýza hlavních komponent (Principal Component Analysis)
<b>PLS</b>	Algoritmus nejmenších čtverců (Partial Least Square)
<b>R6G</b>	Rhodamin 6G
<b>RM</b>	Ramanova mikroskopie
<b>ROC</b>	Operační charakteristika přijímače (Receiver Operating Characteristics)
<b>RS</b>	Ramanova spektroskopie
<b>SERS</b>	Povrchem zesílená Ramanova spektroskopie (Surface Enhanced Raman Spectroscopy)
<b>SVM</b>	Metoda podpůrných vektorů (Support Vector Machine)

---

# Úvod

Ramanovu spektroskopii lze popsat jako nedestruktivní neinvazivní optickou analytickou techniku patřící mezi vibračně-rotační metody. Její velkou výhodou se stala bezesporu minimální příprava vzorků, vysoká rychlost jednotlivých měření, možnost měření vzorku jak pevného, kapalného tak plynného skupenství. Hlavním přínosem je ale skutečnost, že v dnešní době existují přenosné spektroskopy, které mohou být použity v terénu mimo laboratoř. Ramanův jev, který je podstatou RS popsal indický vědec sir Chandrasekhara Venkanta Raman. Ramanova spektroskopie se stala oblíbenou metodou díky možnosti identifikace látek na základě jejich charakteristických spektrálních pásů (fingerprintu) [1].

Své uplatnění si vibrační technika nachází v potravinářství [2], historii [3], forenzní analýze [4], nebo dokonce v lékařské diagnostice [5], kde v dnešní době čím dál častěji nachází své využití. Několik vědeckých skupin se v posledních dvou desetiletích zabývá analýzou klinických vzorků, např. zubů, slin, krve, plazmy, nehtů nebo tkáně. Variabilita jednotlivých vzorků jasně naznačuje širokému užití. Publikované práce vědců se zaměřením na analýzu klinických vzorků se ve svých provedeních liší nejen používaným materiálem, ale také odlišným proměřením vzorků a jejich zpracováním. Autoři článků se zaměřují více na zpracování dat, než na jejich předúpravu, což může zapříčinit menší množství vzorků, které takovou pozornost nevyžadují. Práce, které by se zaměřovaly na možnosti rozdělení jakýchkoliv druhů klinických vzorků podle pohlaví za pomoci RS, byly k roku 2023 publikovány pouze dvě [6], [7]. Žádná z nich ale nepracuje se vzorky plazmy.

Cílem bakalářské práce je naleznout takovou funkční metodu, která by na základě vzorku lidské plazmy proměřené pomocí vibrační techniky, konkrétně Ramanovy spektroskopie, dokázala odhalit ženské nebo mužské pohlaví osoby, k němuž vzorek plazmy patří. Práce se z velké části zaměřuje na předúpravu dat, která je při hledání nejvhodnějších přístupů klíčová. Pokud by bylo možné spektra na základě vhodných algoritmů rozřadit s optimální pravděpodobností úspěšného výsledku, potenciál této techniky by mohl být využit i na jiné složitější parametry. Úkolem této práce není sdělit, jakým způsobem se naměřená spektra liší, ale pokud existuje nějaká šance nalézt algoritmus ukazující podobnost dat na základě stejného pohlaví ze vzorku plazmy.



# Teoretická část

## Kapitola 1

### Ramanova spektroskopie

Ramanova spektroskopie se řadí mezi vibračně-rotální techniky umožňující snadnou interpretaci a vysoce citlivou strukturní identifikaci chemických látek na základě jejich jedinečných vibračních charakteristik (otisků prstů). Patří nejen mezi vibrační, ale také mezi nedestruktivní neinvazivní optické analytické techniky. Od této metody poskytující informace nejen o molekule, ale dokonce i o její struktuře, lze očekávat poměrně rychlou, nedestruktivní analýzu bez nutnosti úpravy vzorku. Zkoumaný materiál navíc může být pevného, kapalného nebo dokonce plynného skupenství. Ramanova spektroskopie se tak stala velmi oblíbenou metodou a nachází uplatnění v řadě oborů od potravinářství, přes průmysl až po medicínu [8], [9].

#### 1.1 Historie metody

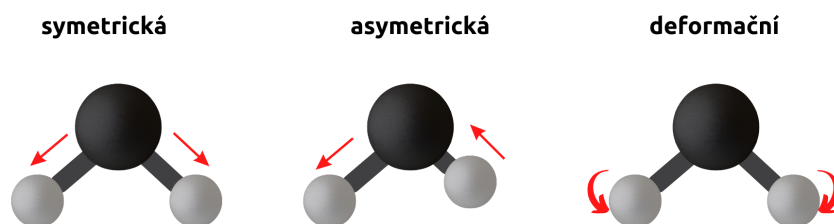
Za objevení Ramanova jevu vděčíme siru Chandrasekhara Venkata Ramanovi, kterému se podařilo experimentálně prokázat fenomén, který už v roce 1923 teoreticky popsal a předpověděl rakouský fyzik Adolf Smekal v článku „Zur Quantentheorie der Dispersion“ [10]. Pokusem se povedl prokázat rozptyl neelastického světla, kdy výstupní záření mělo nepatrně odlišnou vlnovou délku od původního vstupujícího záření, což je zároveň základní myšlenkou jevu, který indický vědec sir Raman v roce 1928 experimentálně dokázal a za svůj objev získal v roce 1930 Nobelovu cenu za fyziku [1], [11].

#### 1.2 Základní teorie

Pro pochopení podstaty Ramanovy spektroskopie a jevu, na kterém je metoda založena, je nejprve nutné si definovat některé základní pojmy.

Na foton, který popisuje kvantum elektromagnetické energie, můžeme pohlížet jako na částici, ale také jako na vlnu. Z tohoto vlnově-částicového dualismu vychází první ze

dvou teorií popisování Ramanova jevu. Světelné částice (fotony), které vznikají při interakci světelného paprsku monochromatického laseru s látkou, mohou být absorbovány, rozptýleny, projít skrz látku a nebo s látkou vůbec nemusí interagovat. Jednotlivé jevy se od sebe liší v závislosti na množství vyzářené energie fotonu. U vibračních metod lze očekávat primárně dva jevy – absorpce, rozptyl. V momentě, kdy energie vyzářeného fotonu odpovídá energii, která je rozdílem mezi excitovaným stavem molekuly a jejím původním stavem, pak může dojít k absorpci, která je typická například pro infračervenou spektroskopii. Druhým příkladem jevu typickým pro tyto metody je bezpochyby rozptyl, který nastává, pokud foton interaguje s molekulou a dochází k rozptylu fotonů. Rozptyl je typický právě pro Ramanovu spektroskopii [1]. Energetické změny, které jsou deteko-

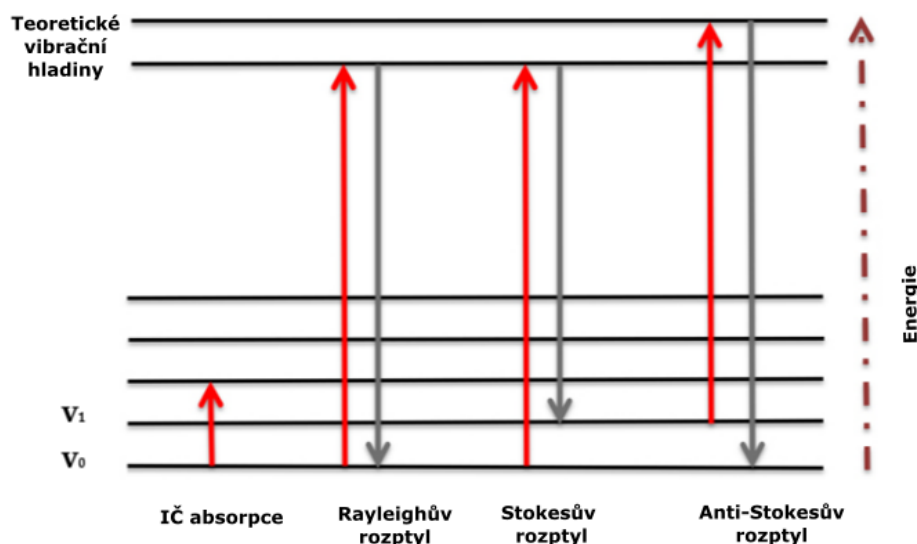


Obrázek 1.1: Typy vibrací molekul

vány při vibrační spektroskopii, jsou změny potřebné k pohybu jader. Jádra vychýlená z osy přilehlých atomů nebo deformace elektronového obalu způsobují vibraci molekuly. K těmto procesům dochází při střetu záření s molekulou. Celkem existují 2 typy vibrací, které jsou zobrazeny na obrázku 1.1 – valenční, kdy se zároveň mění velikost elektronového oblaku, protože nastává změna délky vazby asymetricky nebo symetricky. Druhým typem vibrace je deformační, při které se mění úhel mezi atomy. V případě, že se na rozptylu podílí deformace elektronového oblaku, pak budou fotony rozptýleny s velmi malými změnami frekvence, protože jsou elektrony rozptýlené. Celý tento proces se označuje jako Rayleigho elastický rozptyl. Pokud je však během rozptylu vyvolán pohyb jádra, dojde k přenosu energie buď z dopadajícího fotonu na molekulu, nebo z molekuly na molekulu, v obou případech hovoříme o Ramanově neelastickém rozptylu. Ramanův rozptyl je mnohem slabší než Rayleighův a při přechodu na počáteční hladinu dochází k uvolnění nebo ztrátě energie, která je klíčová pro podstatu Ramanova jevu [1].

Ramanův jev tedy popisuje dva druhy rozptylů – Rayleigho (elastický) a Ramanův (neelastický) rozptyl. V případě neelastického rozptylu je nutno definovat, zda se jedná o anti-Stokesových nebo Stokesových liniích, které jsou ilustrativně zobrazeny na obrázku 1.2. V případě Stokesových linií dojde k excitaci paprsku z  $\nu_0$  do pomyslné vibrační hladiny, ze které se vrací na vyšší vibrační hladinu  $\nu_1$  než byla hladina původní, tedy dojde k vyzáření fotonu o menší energii. V některých případech díky energii uvolněného tepla foton vychází z vyšší vibrační hladiny  $\nu_1$  a z pomyslné energetické hladiny se vrací na základní vibrační hladinu  $\nu_0$ . Vyzářená energie fotonu je tedy větší než energie původní. Smyšlené neboli virtuální energetické hladiny jsou poněkud složitější na představivost, ale díky nim je pochopení podstaty Ramanova jevu mnohem snazší [1], [13].

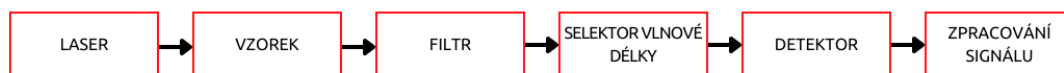
Řada zdrojů srovnává Ramanovu spektroskopii (RS) s infračervenou spektroskopií (IR), tyto dvě metody se navzájem spíše doplňují. Největším rozdílem jsou výběrová pravidla, kterými se obě techniky řídí. Rozptyl Ramanův vyžaduje povinnou změnu polarizace, zatímco IR molekuly jsou aktivní při změně dipólového momentu [14].



Obrázek 1.2: Energie rozptýlených fotonů, upraveno z [12]

### 1.3 Instrumentace Ramanovy spektroskopie

V podkapitole 1.2 byly popsány teoretické principy analytické metody Ramanovy spektroskopie a následující text se bude věnovat instrumentaci této metody, která je názorně ukázána na obrázku 1.3.



Obrázek 1.3: Schéma Ramanovy spektroskopie

Pro získání kvalitních spekter s úzkými pásy s co největší intenzitou vyžaduje technika především kvalitní zdroj záření. Mělo by se jednat o monochromatické záření o jedné vlnové délce, které poskytuje laser. Light Amplification by Stimulated Emission of Radiation představuje název daleko používanější zkratky optického zdroje. Vzhledem k malé intenzitě Ramanova rozptylu je nutné zvolit co možná nejvýkonnější zdroj záření, který putuje přes řadu zrcadel na vzorek. V moderních přístrojích však lze získat rozpoznatelná spektra i s nízkovýkonnými lasery. Oddělení vlnově posunutého Ramanova rozptylu, který je mnohem intenzivnější od ostatního shromážděného světla, je možné za pomoci filtrů. Přes řadu filtrů a zrcadel se signál dostává na selektor vlnové délky, odkud putuje optickými kabely na detektor s cílem co nejefektivnějšího zobrazení zachyceného světla. Detektor je v posledním kroku připojen k počítači, který za pomoci vhodných softwarů dokáže signál zobrazit ve formě spekter. Schématem instrumentace 1.3 se řídí většina Ramanových spektroskopů, další techniky od něj odvozené se ale mohou lehce lišit [1], [15].

Trend minimalizovat svět zasáhl i Ramanovu spektroskopii. Ještě před několika lety jsme znali pouze veliké přístroje umístěné v temných komorách bez možnosti promě-

řování vzorků mimo místnosti laboratoře. Dnes se již běžně setkáváme s přenosnými spektrometry, které se díky svojí velikosti vejdou i do kapsy uživatele [1].

## 1.4 Ramanovo spektrum

Grafickým záznamem z RS je Ramanovo spektrum, které je zobrazováno jako graf závislosti intenzity na Ramanovském posunu, který je možný definovat jako

$$\Delta\tilde{\nu} = \tilde{\nu}_0 \pm \tilde{\nu} \text{ (cm}^{-1}\text{)}, \quad (1.1)$$

kde  $\tilde{\nu}_0$  představuje frekvenci Rayleigha rozptylu a  $\tilde{\nu}$  potom frekvenci Stokesovu ( $\tilde{\nu}_0 + \tilde{\nu}$ ) nebo anti-Stokesovu ( $\tilde{\nu}_0 - \tilde{\nu}$ ) [13]. Podle ustanovení IUPAC by se energie emitovaného záření měla označovat jako  $\Delta \text{ cm}^{-1}$ , ve většině vědeckých publikací se ale uvádí označení osy x jako  $\text{cm}^{-1}$ . Rozsah Ramanova spektra se nejběžněji pohybuje v rozmezí od  $200 \text{ cm}^{-1}$  do přibližně  $3000 \text{ cm}^{-1}$ , kde se projevuje nejvíce molekul. Všechny pásy v Ramanově spektru stoupají od základní linie proti lineární stupnici. Tato stupnice se mění od přístroje k přístroji a nelze ji snadno použít k přímému kvantitativnímu měření. Intenzity v Ramanově spektru závisí na povaze vibrací, které jsou předmětem zkoumání, ale zároveň ji ovlivňuje řada faktorů, jako je výkon laseru použitého při excitaci. Samotné spektrum ale v některých případech vyžaduje řadu procesů, díky kterým můžeme snadněji data mezi sebou porovnávat. V tomto případě upravování dat nepředstavuje nic nelegálního, ale již dlouho zavedené postupy, kterými se vědci řídí. Mezi základní kroky úpravy dat se řadí například úpravu šumu, korekci spektrálního pozadí nebo celkovou normalizaci spekter. Jednotlivé pásy mohou být identifikovatelné za pomoci přístupných i komerčně dostupných softwarů doplněné o knihovny. Spektra poté mohou posloužit pro kvalitativní účely [1].



## Kapitola 2

# Ramanova mikroskopie

Následující kapitola zahrnuje popis jedné ze speciálních technik Ramanovy spektroskopie, kterou je Ramanova mikroskopie (RM). Čtenář se seznámí se základními pilíři instrumentace a pronikne i do bohatého využití této analytické techniky.

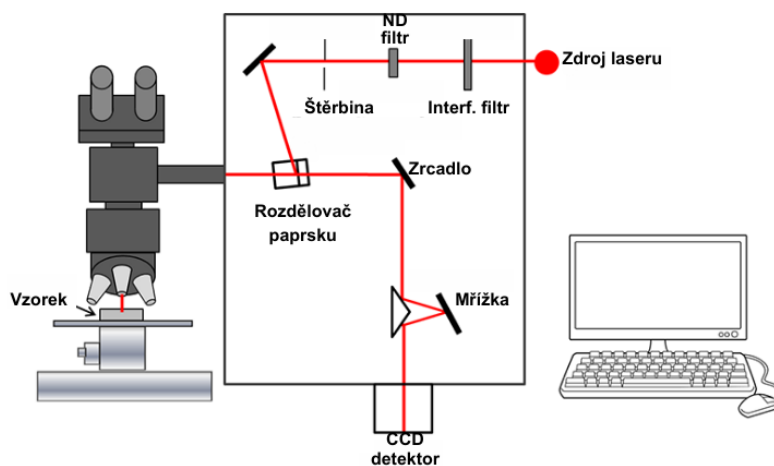
### 2.1 Ramanova mikroskopie

Po popsání Ramanova jevu v roce 1930 měli vědci od spektroskopické metody velká očekávání. Jedna z mála nevýhod, která se objevila u Ramanovy spektroskopie, byly problémy s fluorescencí. Vzhledem k tomu, že některé látky při excitaci fluoreskují, a to s poměrně velkou intenzitou, může tak docházet k překryvům Ramanova spektra, což je samozřejmě nežádoucí efekt. Ramanova mikroskopie vychází z podstaty Ramanovy spektrometrie a byla poprvé vyzkoušena v sedmdesátých letech minulého století a od té doby se jen a jen rozvíjí. Největší limitací pro tuto vibrační techniku představuje nízká intenzita Ramanova signálu. Velká pozornost při sestavování techniky tedy musí připadat na přední optickou konfiguraci, aby nedocházelo ke ztrátám rozptýleného světla, ale aby veškeré paprsky tvořící následný signál byly zachycovány na detektoru, který musí být také vysoce citlivý. Snahou je také poskytnout samostatnou dráhu pro zachycení excitovaného a rozptýleného záření. Rovněž použitý substrát na laboratorním sklíčku musí vykazovat co nejmenší známky fluorescence a Ramanova signálu v požadovaném rozsahu měřeného vzorku. Vrstva substrátu by měla být leštěná, aby mohlo dojít k viditelnému pozorování. Častým materiálem je fluorid vápenatý, oxid manganatý, v některých případech je možné použít pouze sklo [13].

### 2.2 Instrumentace metody

K získání Ramanova spektra je zapotřebí monochromatického záření o jedné vlnové délce, tedy laseru. Pokud má přístroj alternativu připojení více laserů, získává uživatel mnohem lepší možnost pro lepší výsledky měření. Výběr laseru se liší na základě vlastností vzorku, chemickou i biologickou povahou materiálu a jiné. Zdroj vyzáří paprsek, který přes řadu filtrů umístěných přímo v mikroskopu, které odstraňují silné Rayleighovo záření, putuje skrz štěrbinu a několik zrcadel na vzorek viz obrázek 2.1. Vzorek umístěný na stolku mikroskopu je zaostřen pomocí optického zobrazení se spodním nebo horním

nasvícením. V momentě, kdy je vzorek zafokusován skleněnými čočkami, nasvícení je umístěno do polohy vypnuto a na analyt míří pouze fotony ze zdroje. Měření tak probíhá v pokud možno co největší tmě, aby nedocházelo k rušení signálu. Rozptýlené světlo je zachycováno objektivem a dále již putuje na detektor, který je spojený s počítačem a programem umožňujícím zobrazit Ramanovo spektrum. Výhodou Ramanova mikroskopu je především prováděné odfiltrování přímo uvnitř mikroskopu, lepší fokusace vzorku, možnosti použití i malého množství vzorku (teoretické prostorové rozlišení je  $1 \mu\text{m}$ ), ale také umožnění měření několika vzorků najednou při spojení přístroje s vhodným počítačovým softwarem. Jako velkou výhodou Ramanovy mikroskopie se může jevit i použití metody bez větší znalosti teoretického základu Ramanova jevu, pochopení teorie ale uživateli pomůže zejména při zpracování dat [1], [13].



Obrázek 2.1: Instrumentace Ramanovy mikroskopie

## 2.3 Obecné využití

Přístroje fungující na principu Ramanova jevu se objevují v laboratořích čím dál častěji. V předešlé kapitole byly zmíněny aplikace Ramanovy spektrometrie, která se ukázala jako velice univerzální metodou vzhledem k její funkčnosti. RM je speciální metoda, která se ale v posledních letech dostává více a více do popředí. Propojení spektrometru s mikroskopem ukázalo výhody v minimální přípravě vzorků, malého množství analyzovaného materiálu, zesílení Ramanova signálu a nebo dokonce omezení negativního jevu, jakým může například být fluorescence. Nejen díky těmto vlastnostem si metoda své využití nachází v biologických, chemických i medicínských aplikacích. Metoda se ukázala jako vhodná v kombinaci s dalšími zobrazovacími technikami pro kompletní vizualizaci a charakterizaci i rakovinotvorných buněk. Ramanovo zobrazování by mohlo být přínosem pro chirurgy, protože by umožnilo úplnější vizualizaci vzorků nádorů, nebo by mohlo sehrát určitou roli ve studii přenosu léčiv [16]. Významných pokroků bylo dosaženo v archeologii dosaženo díky využití Ramanovy mikroskopie a došlo k odhalení oblasti vyžadující další výzkum, jako je vytváření rozsáhlejších a spolehlivějších spektrálních knihoven a překonávání překážek při analýze některých tříd historických materiálů [14].

## Kapitola 3

# Využití v praxi

V posledním desetiletí se Ramanova spektroskopie ukazuje jako mocný nástroj pro řadu analýz a nachází uplatnění v mnoha odvětvích. Na základě charakteristických spektrálních vzorků (fingerprint) je možné používat metodu k poskytování informací o chemické struktuře a fyzikálních formách, ale také ke kvantitativnímu a semikvantitativnímu určení množství látky. Velkou výhodou vibrační techniky je především možnost měření látek s různými fyzikálními vlastnostmi jako například v pevném, kapalném nebo plynném stavu, možnost měření při různých teplotních podmínkách, mikroskopické částice nebo povrchové vrstvy [1].

### 3.1 Potravinářství

V roce 2012 se mnoha obyvatel České Republiky dotkla tzv. Methanolová kauza, kdy bylo zjištěno v několika desítkách případech nadměrné množství methanolu v alkoholových destilátech, což v mnoha případech došlo k oslepení nebo bohužel až k fatálním následkům, tedy smrti. Při této kauze byla využita Ramanova spektroskopie díky možnosti proměřovat látky skrz skleněné obaly. Běžné metody detekce biologických, chemických a fyzikálních kontaminantů v potravinách jsou pracné, drahé, časově náročné a často také destruktivní. Ramanova spektroskopie se jeví jako vhodnou metodou pro zabezpečování kvality potravinových produktů a do budoucna by tak mohla urychlit a usnadnit analýzy zaměřované na pesticidy, toxiny nebo bakterie a viry často v jídle obsažené [2], [17].

### 3.2 Historie

Vzhledem k tomu, že je metoda považována za nedestruktivní typ analýzy, představuje vhodného kandidáta pro analyzování vzorků umělecko-historických děl, archeologických nálezů nebo posuzování historických budov. Nedochozí totiž ke znehodnocení materiálu. Výhodou v aplikacích tohoto druhu může být i malá spotřeba daného vzorku bez nutnosti přípravy. V případě uměleckých obrazů může posloužit například jako nástroj pro charakteristiku použitých barviv nebo jejich degradaci za účelem získání potřebných informací o látce, které mohou využít například restaurátoři cenných děl. Ramanova mikroskopie přispěla také z hlediska archeologických výzkumů, kdy se povedlo s určitou pravděpodobností stanovit pohlaví nalezených těl pouze za pomoci vzorků ze

zubů, nebo v případě používaného léku starověkého světa Lykionu nalezeného v 15 publikovaných keramických nádobách [3], [18].

### 3.3 Forenzní analýza

Jedním z oborů, ke kterému Ramanova spektroskopie výrazně přispěla, je forenzní věda. RS se ukázala být cenným nástrojem pro analýzu různých typů forenzních důkazů, například k analýze tělních tekutin, lidských tkání, kostí, zbytků střelných zbraní, vláken a vlasů. Vyvinutý postup v článku [19] vykazuje velký potenciál pro rozlišení kuřáků a nekuřáků na základě analýzy suchých stop ústní tekutiny. RS může být přínosem pro kriminalistické obory z několika důvodů: umožňuje získat rychlé, nestranné a statisticky spolehlivé výsledky v reálném čase. S metodou RS je možné se setkat při letištních kontrolách, kdy odhaluje i nepatrné množství nebezpečných a zakázaných látek [4], [20].

### 3.4 Lékařská diagnostika

Mezi aplikace Ramanovy spektroskopie v přírodních vědách patří kvantifikace biomolekul, hyperspektrální molekulární zobrazování buněk a tkání, lékařská diagnostika a další [5]. Lékařská diagnostika je velkým tématem pro řadu vědců, čemuž nasvědčuje pouhý výčet ze zajímavých prací z posledních let. Příklady prací jsou jasným důkazem, jak moc aktuální téma to je i pro využití RS. Studie [21] ukázala, že Ramanova spektroskopie je technika, kterou lze využít k biochemickému rozlišení leukemií a jiných typů rakoviny ve vzorcích séra dětí a dospívajících, a to přibližně s 67% pravděpodobností. V článku [20] se hovoří o testování RS jako vhodného nástroje pro diagnostiku různých typů stádií jaterních onemocnění. Ukazují se také možnosti využití RS jako pomocného nástroje v řadě diagnostických analýzách sloužící k určení například Parkinsonové choroby [22]. Po vypuknutí celosvětové pandemie v souvislosti s onemocněním Covid-19, bylo nutné odlehčit laboratořím, které prováděli testy na přítomnost koronaviru SARS-CoV-2. V roce 2020 byl publikován článek [23], ve kterém bylo popsáno možné využití SERS pro alternativní spolehlivou diagnostiku detekce SARS-CoV-2. Za podstatné vědecké pracovníky považují možnost využití spektroskopické metody při možném budoucím vypuknutí virové pandemie.

## Kapitola 4

# Metody analýzy klinických vzorků pomocí RS

Text této kapitoly se zabývá jednotlivými metodami analýzy klinických vzorků, které posloužily jako inspirace pro téma bakalářské práce. Jednotlivé články ukazují úskalí a limity této techniky a zároveň zpracování výsledků, které se u jednotlivých prací liší. Postupně zde bude ukázáno, proč a na základě jakých znalostí byly v experimentální části použity právě tyto postupy.

Z kapitoly o využití vibračních metod v praxi vyplynulo, že si své uplatnění RS nachází snadno téměř ve všech odvětvích. V posledních letech se čím dál častěji objevují publikace zaměřující se na analýzu klinických vzorků. Cílem bývá nejčastěji rozpoznání nemocí, ať už zhoubných, jako je rakovina, nebo nezhoubných. Klinické studie nemusí využívat metodu vždy jako hlavní nástroj při odhalování nebo analýze nemocí, často mohou prokázat své služby i jako pomocný prostředek a to nejen v oboru lékařství. Studie o možnostech analýz klinických vzorků se liší v mnoha aspektech. Využívají různých materiálů o různém počtu vzorků, pracují při různých experimentálních podmínkách a také pracují s různými metodami zpracování naměřených dat.

### 4.1 Materiál, množství vzorků

Analýzou klinických vzorků, mezi které patří vzorky tělních tekutin, zubů, lidské tkáně aj., se v minulosti zabývalo několik vědeckých skupin, které při svých výzkumech využívali jednu z vibračních metod – Ramanovu spektroskopii. Například v článku [24] z roku 2022 se autoři pokusili ukázat RS jako nástroj při diagnóze rakoviny hlavy a krku. Práce vede k diskuzi, jak přínosné mohou být vzorky krve a slin proměřené touto vibrační technikou. Celkem bylo odebráno 53 vzorků pacientům s cílem klasifikovat onemocnění. Ukázalo se, že kombinace výsledků získaných z krve a slin může poskytnout vyšší efektivitu a kvalitu výsledků až na 96 %, zatímco samotné vzorky plazmy a slin vykazují úspěšnost mezi 70 až 90 %

Autoři článku [20] popisují Ramanovu spektroskopii jako metodu vhodnou pro studování biomedicinských vzorků, ale upozorňují na neprokázání účinného využití v klinické diagnostice. Poměrně velkým rozdílem od dvou předešlých článků je počet vzorků. V této studii bylo 234 vzorků séra ze souboru pacientů s předem stanovenou diagnózou

různými stádii jaterních onemocnění. Vzorky byly proměřeny při excitační vlnové délce 780 nm za pomoci vibrační techniky Ramanovy spektroskopie. Takovéto druhy onemocnění si žádají časté laboratorní testy, které se provádí jednotlivě na různé biomarkery (indikátory určitého biologického stavu) samostatně. Cílem studie bylo zhodnotit využitelnost Ramanovy spektroskopie jako diagnostický nástroj pro klinické vzorky v porovnání s rutinními laboratorními diagnostikami. Velkou část výzkumné práce je věnovaná zařazení specifických biomarkerů pro jaterní onemocnění. Výsledky ukázaly, že Ramanova spektroskopie má vysoký potenciál pro využití v lékařské laboratorní diagnostice ke kvantifikaci více biomarkerů současně. Opět jako již v předešlém článku je zde určita variabilita pacientů, kterou způsobuje odlišné stádium onemocnění [20]. V článku [25] byla například RS použita k bližšímu popisu struktury zubů v oblasti stomatologie. Spektrální pásy spojené s organickými a anorganickými částmi skloviny a dentinu (mineralizovaná pojivová tkáň) by mohly být užitečné při předpovídání časného vzniku kazů.

## 4.2 Proměření

Výše uvedené příklady ukazují především velikou škálu využití metody RS při zpracování různých klinických vzorků. Odlišnost vzorků nabízí také využití různého proměření. Například vzorky krevního séra v článku [20] byly proměřeny při vlnové délce 780 nm, v další studii [24] při vlnové délce 785 nm. Ale naopak v článku [26] bylo zhodnoceno využití vlnové délky 785 nm jako nevhodné při použití podložního skla, a proto byla zvolena vlnová délka 532 nm. Výhodou tohoto uspořádání je bezpochybně to, že poskytuje vysoce kvalitní a konzistentní Ramanova spektra z objemů vzorků pouhých 1  $\mu$ l, což značně usnadňuje použití RS při proměření klinických vzorků, kterých se odebírá malé množství, tudíž je možné analýzy opakovat, nebo ponechat zbylý objem pro jiné techniky. Laser s nižší vlnovou délkou při proměrování na použitém substrátu  $\text{CaF}_2$  byl doporučen i v článku [27].

## 4.3 Zpracování

Předtím, než je možné aplikovat metody pro zpracování dat, je nutné naměřená data vizuálně projít a přistoupit k nutným předúpravám spekter, což je nezbytným krokem pro získání výsledků s dostatečnou přesností. Mezi takové kroky patří korekce šumu, odstranění spektrálního pozadí, odebrání nepotřebných částí spekter nebo normalizace dat.

Vzorky krevního séra ve studii [24] byly testovány ve dvou fázích, kdy v první fázi byla pozornost přenesena na obecné testy, které hodnotily použití vibrační techniky na klinické vzorky, což přineslo informaci především o šumu a obsahu látek. Dalším krokem byla korelace dat se známými parametry, které odráží možnost využití Ramanovy spektroskopie v klinické diagnostice. Důležitými kroky pro obě fáze ale byly filtrace pozadí a korelační analýza. Zároveň tato studie zdůrazňuje, že je třeba zabránit nadměrnému přizpůsobení a používat k validaci nezávislá data.

Již zmíněný článek o možnosti rozpoznání pohlaví ze vzorku zubů je hlavně zajímavý svým zpracováním. K statistickému zpracování byla vybrána metoda analýzy hlavních

komponent PCA, která sloužila k dekorelaci dat. Práce byla založena na 1 až 20 komponentách, což značí velkou diverzibilitu vzorků. Pro další zpracování byly vybrány dva klasifikační algoritmy

1. klasifikace dat pomocí metody podpůrných vektorů (SVM),
2. umělé neuronové sítě (ANN),

kdy obě metody vykazovaly podobné výsledky, které se pohybovaly okolo 80 %. Výsledky tedy ukázaly, že záleží nejen na typu zubu, ze kterého vzorek pochází, ale také na místě zaznamenaného spektra [7].

#### 4.4 Studie zaměřené na rozpoznání pohlaví

Prozatím existují v oblasti studie Ramanovy spektroskopie využité k rozpoznání pohlaví za pomoci klinických vzorků pouze dva publikované články svého druhu. První publikace (*A novel method for human gender classification using Raman spectroscopy of fingernail clippings*) z roku 2008 využila pro svůj výzkum vzorky nehtů proměřených na RS. Celkem bylo proměřeno 80 různých vzorků nehtů (40 z nich patřilo mužům a 40 patřilo ženám). Všichni dárči materiálu byli ve věku 24 až 50 let. Ke klasifikaci dat byl využit algoritmus analýzy hlavních komponent (PCA) a podpůrných vektorových strojů (SVM). Tento kombinovaný algoritmus poskytoval klasifikační přesnost až 90 %. Dosažený výsledek ukazuje možnost využití metody jako alternativní rychlý nástroj pro identifikaci lidského pohlaví například ve forenzních aplikacích [6].

Autoři druhého článku (*Possibility of Human Gender Recognition Using Raman Spectra of Teeth z roku 2021*), který se také zabýval rozlišením pohlaví za pomoci RS, využívali jako materiál lidské zuby. Autoři se soustředí na rozpoznání pohlaví z kousku chrupu hlavně pro archeologické účely. Proces identifikace archeologických vykopávek a pozůstatků představuje řadu procesů, které jsou nesmírně komplikované a začínají právě rozpoznáním pohlaví osoby, věku, postavy. Z výše uvedeného vyplývá, že je pro tento obor velice zásadní tyto parametry rozlišit pro další zkoumání. Často se při objevování ostatků povede objevit nejtvrděší a jeden z nejdéle trvajících materiálu lidského organismu nacházející se v dutině ústní, kterým jsou zuby, na které se tato studie zaměřuje. Pro 55 vzorků zubů patřící věkové skupině 11 – 76 let byla použita konkrétně metoda FT-Raman spektroskopie. Zpracování dat probíhalo opět za pomoci PCA analýzy, SVM metody a mimo jiné i s použitím metody umělé neuronové sítě (ANN). Závěrem práce bylo zjištěno, že došlo ke klasifikační přesnosti vyšší než 90 %, ale výsledky se lišily od typu zubu i od místa jeho proměření [7].

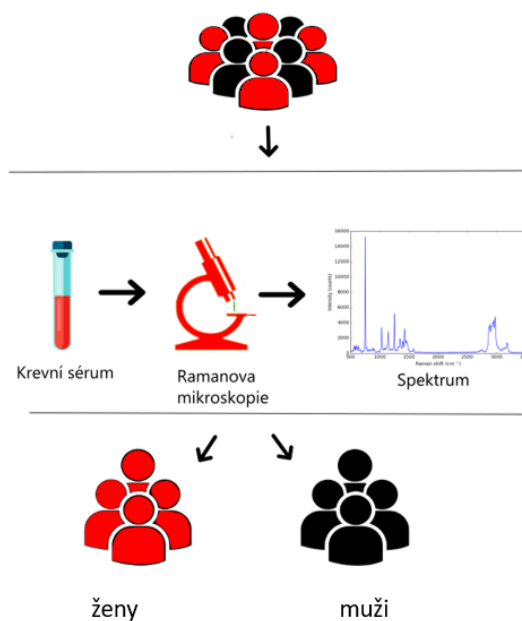




# Experimentální část

## Cíl práce

- Naleznout metodu pro efektivní a spolehlivé proměření dat
- Proměřit vzorky pomocí Ramanovy spektroskopie
- Na základě výsledků porovnat jednotlivé přístupy



Obrázek 4.1: Grafické znázornění podstaty bakalářské práce



## Kapitola 5

# Vlastní měření spekter

V předchozích kapitolách byl čtenář uveden do kontextu metody Ramanovy spektroskopie. K pochopení podstaty uvedené vibrační techniky pomohlo popsání základní teorie, instrumentace RS ve spojení s běžným využitím v praxi. V poslední kapitole teoretické části byly citovány články, které posloužily jako inspirace pro výběr tématu odborné práce. Z obecných informací bude nyní přistoupeno k praktickému provedení práce, kde bude popsáno, jaké konkrétní vybavení a materiál byly v experimentu použity. Při popisu bude kladen důraz na objasnění výběru vybavení.

### 5.1 Studium přístupu měření

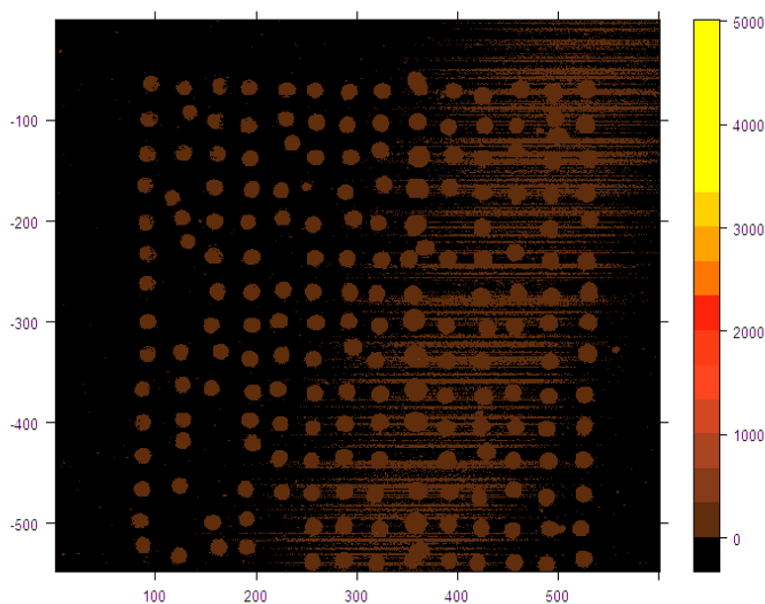
Předtím, než bylo přistoupeno k samotnému experimentu na reálných vzorcích, bylo nutné se seznámit s mechanismy metody. Proto byla vybrána látka, která by pro první pokusy byla vhodná a na které by bylo dobré jednotlivé postupy vyzkoušet. Rhodamin 6G (často označován pouze jako R6G) je látka patřící do skupiny rhodaminových vysoce fluorescenčních barviv, která je navíc mísitelná s vodou, takže umožňuje snadné ředění roztoku a je vhodná pro měření na RS. Na proměření barviva byly otestovány předpoklady Ramanovy spektroskopie pro následující analýzu. Od spektroskopické metody totiž byla vyžadována co nejmenší diverzita mezi naměřenými spektry jednotlivých kapek. Data by se od sebe neměla lišit ani v závislosti na umístění měřeného vzorku na laboratorním sklíčku, ani na zvoleném substrátu. Intenzita a vlnová délka laseru by měla být zvolena přiměřeně k vlastnostem materiálu, zkoumaná byla také délka jednotlivých skenování při analýze. Pro experiment bylo nutné vyzkoušet vhodný substrát nanesený na laboratorním sklíčku, aby vykazoval co nejlepší homogenitu, neposkytoval významný signál a nefluoreskoval. Pro první část experimentu byla zvolena vrstva zlata, kdy se však nejednalo o nanomateriálovou vrstvu s povrchovým plasmonem, který by byl využitelný pro metodu SERS (Povrchem zesílenou Ramanovu spektroskopii). Vzhledem k tomu, že se na použitý substrát kladl důraz především z hlediska co největší homogenity vrstvy k zajištění požadované technické opakovatelnosti a reprodukovatelnosti měření, nebyl důvod o možnosti metody SERS vůbec uvažovat.

Nevýhodou skla potaženého vrstvičkou zlata je zajisté vyšší cena, ale také snazší mechanické poškození vrstvy. Proto byl vyzkoušen další substrát, kterým byl fluorid vápenatý. Vykazoval téměř stejné vlastnosti, tudíž mohl být také použit při další analýze. Naměřené výsledky neshledaly výrazné rozdíly v měření při použití jednoho z těchto

dvou substrátů. Z hlediska uchování nanesených kapek na sklíčku bylo finančně méně náročnější pro další analýzy využít materiál  $\text{CaF}_2$  [28].

Po proměření roztoků na laboratorním sklíčku s povrchem zlata o koncentraci v rozmezí  $10^{-6}$  až  $10^{-3}$  mol/l bylo zjištěno, že nejlepší intenzity dosahuje R6G při koncentraci  $10^{-6}$  mol/l. Zhruba  $5 \mu\text{l}$  R6G o koncentraci  $10^{-6}$  mol/l bylo pipetováno automatickou pipetou do LDV desky 384 a následně pomocí přístroje Labcyte Echo 550 Liquid handler přeneseno na pozlacené podložní sklíčko do matice  $15 \times 15$  vzorků. Tento přístroj se běžně používá k nanášení vzorků využívající akustické vymrštění kapek, která ultrazvukovou energii soustředí na kapalinu. Inovativní technologie automatizované manipulace s kapalinami se využívá především v oblasti biologických výzkumů.

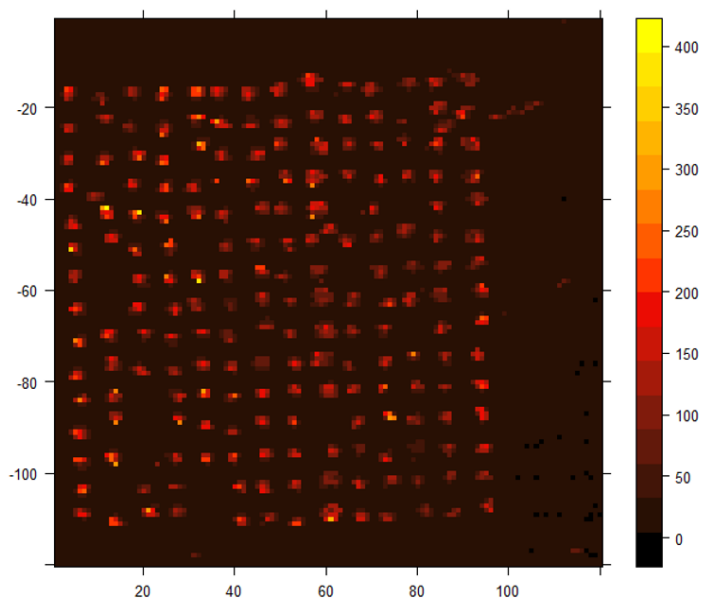
Ramanův mikroskop Witec Alpha 300R+ byl použit v propojení se zdrojem záření o vlnové délce 532 nm s intenzitou laseru na vzorku 5 mW. Délka scanu u první matice vzorků o velikosti zobrazované oblasti  $12 \times 12$  mm s rozlišením  $600 \times 600$  pixelů byla 1 s, ale následně byl čas zkrácen na 0,05 s a rozlišení zmenšeno na  $150 \times 150$  pixelů. Mikroskop pracoval v rozlišení: 20 až  $80 \mu\text{m}$ .



Obrázek 5.1: Ramanská mapa průměrné Ramanské intenzity

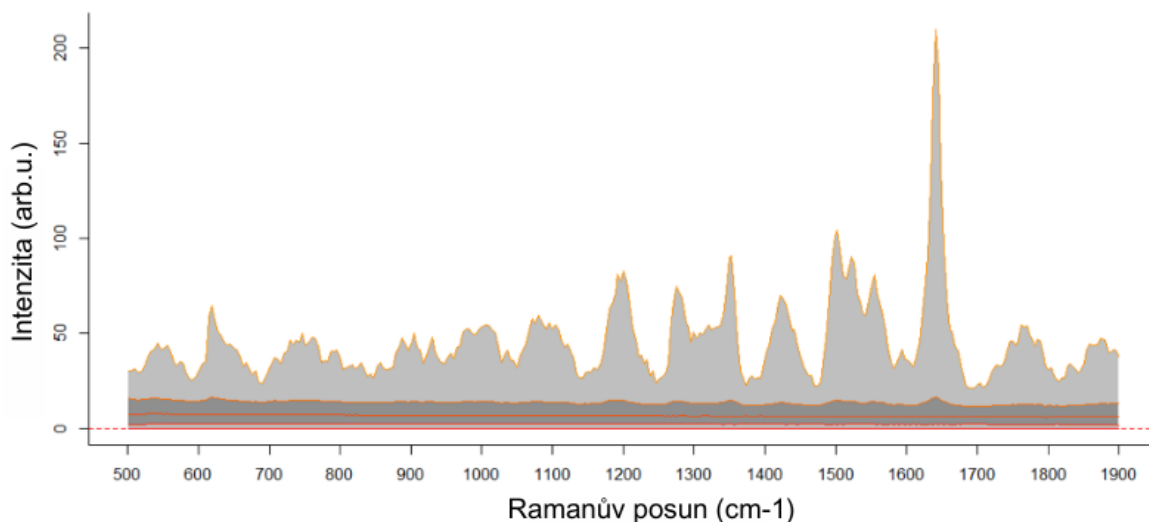
Problémů při urychlení celé analýzy se ukázalo hned několik. Prozatím nebyla vyvinuta metoda pro zpracování výsledné mapy obsahující matici vzorků. Metoda, která by dokázala matici vzorků proměřit a z jednotlivých dat vyjmout spektrum jednoho vzorku, ke kterému by ještě přiřadila čtyřmístný kód patřící konkrétní osobě, zatím nebyla vyvinuta a svojí náročností by přesahovala rámec této bakalářské práce. Bohužel se neukázala téměř žádná kontrola nad měřenými částmi kapek, tedy v jakém konkrétním místě by měla být kapka změřena, vzhledem k nesystematickému usychání kapky. Proto musela být původně zamýšlená metoda nanášení vzorků za pomoci přístroje Labcyte Echo a automatického proměřování za pomoci Ramanských map zamítnuta.

Na obrázku 5.2 je na první pohled zřejmé, že matice o velikosti  $15 \times 15$  vzorků zobrazená v Ramanské mapě ukazuje posuny a rozmytí jednotlivě nanesených kapek R6G, což rovněž ukazuje obrázek Ramanské mapy 5.2 zobrazující univariální analýzu pro



Obrázek 5.2: Univariantní analýza Ramanské mapy

spektrální pás R6G při  $1642\text{ cm}^{-1}$ . Poměrně vysoká chybovost 13 % při nanášení vzorků na podložní sklíčko a smývání kapek do sebe, která by při použití reálných vzorků znemožňovala celou analýzu především pro nemožnost následné identifikace vzorků, čímž by znemožnil celou analýzu. Výše popsané problémy ukazují jasné důvody, proč tento přístup nanášení vzorků není vhodný pro následné použití v této bakalářské práci.

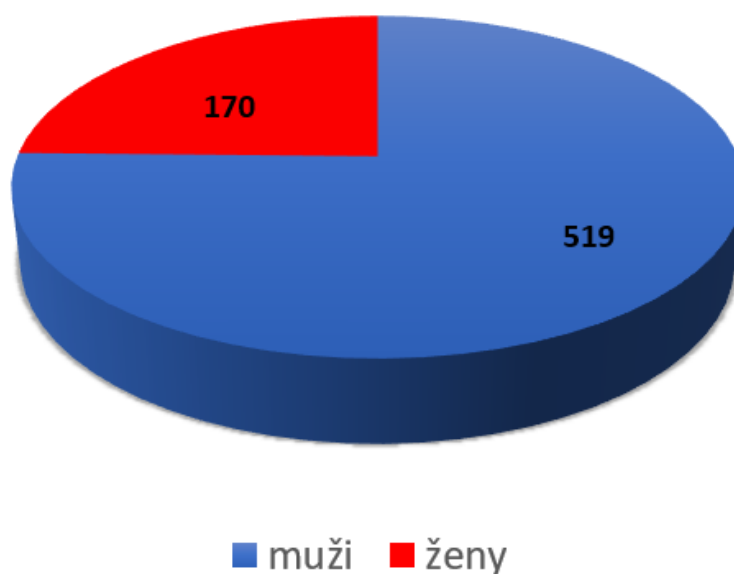


Obrázek 5.3: Průměrná spektra na Ramanské mapě

Jednotlivá spektra byla upravena korekcí pozadí odečtením polynomu 3. stupně, což bylo optimalizováno na základě naměřených dat a za pomoci normalizace spekter podle nejvyššího spektrálního pásu. Průměrné Ramanské spektrum, které je zobrazeno na obrázku 5.3, ukazuje dostatečně velký rozdíl mezi naměřenými vzorky R6G (světle šedé pásy) a spektrem substrátu (tmavě šedé pozadí označené oranžovou linií). Rozdílných spekter by se dalo využít v případě Ramanských map a jejich analýzy.

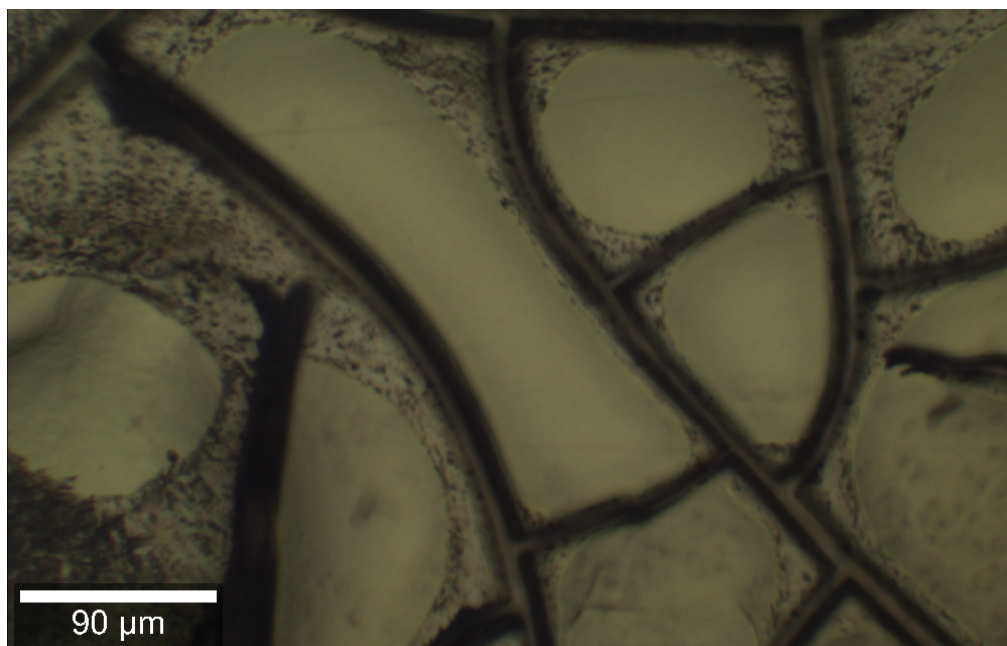
## 5.2 Měření spekter reálných vzorků plazmy

Po seznámení se s metodou RS za pomoci proměřování materiálu R6G, bylo možné přistoupit k proměření reálných vzorků. Jako hlavním materiálem pro experiment bylo vybráno krevní sérum, konkrétně plazma. Vzorky byly odebrány do mikrozkušavek Eppendorf a v boxech uloženy v mrazničce. Z celkového počtu 1103 vzorků plazmy zdravých pacientů bylo proměřeno 689 jedinců, z toho 519 zástupců mužského pohlaví a 170 zástupců ženského pohlaví. Při proměření přibližně poloviny připraveného materiálu bylo zjištěno, že je mnohem víc vzorků patřících mužům než ženám. Zvýšení procenta žen bylo možné zajistit pouze tak, že ze zbývajících zkumavek byly vybrány pouze ty, které patřily ženám. Bylo tedy zbytečné pipetovat další vzorky mužů. Na výsečovém grafu 5.4 je zobrazeno červenou barvou zastoupení žen, které tvořilo 25 %, a zbylých 75 % tvořila většina zbarvená modrou barvou, tedy muži. Kapka biologického materiálu o objemu přibližně 2  $\mu\text{l}$  byla nanášena na sklíčko s povrchem fluoridu vápenatého. Tento materiál byl vybrán především ze dvou hlavních důvodů. Zaprvé nefluoreskuje a za druhé nepřináší nijak zásadní Ramanovský signál, který by rušil naměřená spektra.



Obrázek 5.4: Zastoupení mužů a žen v analýze

Vzorky byly nanášeny pomocí automatické pipety dostatečně daleko od sebe, aby nedocházelo ke smývání více kapek dohromady. Po zaschnutí séra došlo k uskladnění vzorků do lednice tak, aby nedošlo k jejich destrukci. Nanesená plazma vykazovala dva problémy při uskladňování. Za prvé po uložení sklíčka s nanesenými vzorky do úložného boxu docházelo k opadávání zaschlých kapek, tudíž některé vzorky musely být nanášeny vícekrát. Za druhé vzhledem k vysychání kapek docházelo k popraskání jejich povrchu, které je viditelné na obrázku 5.5. I z tohoto důvodu bylo nutné zobrazení naneseného materiálu na sklíčko a zaměření paprsku na místo mimo tmavé čáry, kde bylo možné naměřit signál pro Ramanovo spektrum. Po proměření byla všechna spektra otevřena v programu SpectraGryph 1.2, kde byla podrobena vizuální kontrole a spektra s nižší intenzitou, vyšším šumem nebo v případě, že vykazovala větší intenzitu fluorescence byla následně přeměřena. Je nutné podotknout, že spektra vykazovala poměrně vysokou kvalitu, takže vzorků, které by byly přeměřovány bylo v řádu desítek.



Obrázek 5.5: Kapka plazmy

Ramanův mikroskop značky Witec Alpha 300 R+ ve spojení s laserem o vlnové délce 532 nm a intenzitou záření na vzorek 5 mW byly využity k proměření vzorků pomocí Ramanovy mikroskopie s rozlišením objektivu 20 až 80  $\mu\text{m}$ . Měření probíhala v temné místnosti, aby nedocházelo k narušení měřených spekter, kdy jednotlivé scany trvaly 8 s. Při těchto podmínkách a parametrech shrnutých v tabulce 5.1 byly proměřeny všechny vzorky a následně byly zpracovány v programu Spectragryph 1.2.15 a pomocí softwaru RStudio v jazyku R verze 4.2.2.

Tabulka 5.1: Technické parametry měření

Parametry měření	
materiál	Rhodamin 6G krevní sérum- plazma
Ramanův spektrometr	Witec Alpha 300R+
laser	532 nm
intenzita laseru na vzorku	5 mW
délka scanu	8 s
rozlišení	20 až 80 $\mu\text{m}$





## Kapitola 6

# Spektrální analýza naměřených dat

Úpravy naměřených dat jsou běžným krokem při spektrálních analýzách dat a je nutné k nim přistupovat zodpovědně a zabývat se správným výběrem použitých metod k tomu určených. Proto bude v následující kapitole popsáno, které knihovny a funkce s jejich modifikacemi byly při analýze použity. Veškeré úpravy a následné analýzy byly prováděny v programovacím jazyce R.

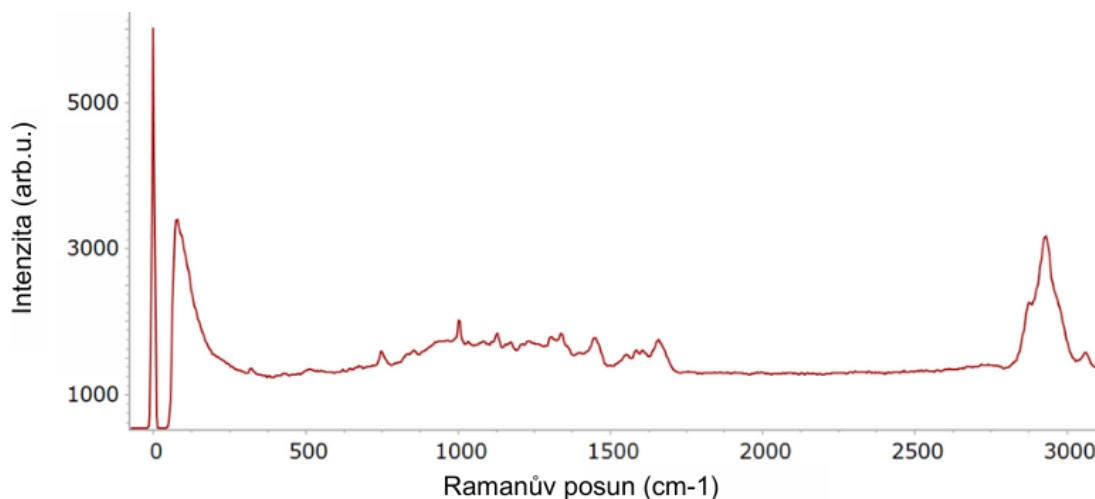
ChemoSpec je soubor funkcí pro výzkumnou analýzu spektrálních dat umožňující zpracování dat z Ramanova záření, infračerveného záření (IR), rentgenova záření, nukleární magnetické rezonance (NMR) a dalších podobných typů spektroskopie. Slouží především jako uživatelsky příjemný nástroj pro chemické analýzy shrnující matematické a statistické funkce. Knihovna ChemoSpec je také vhodná pro zpracování robustnějších souborů s větším objemem dat. Výhodou odstranění přebytečných dat je zrychlení procesu, ale také to vede k lepším dosaženým výsledkům. Knihovna funguje na poměrně jednoduchém principu, kdy z nijak nezpracovaných dat nejprve odebere data, která nejsou vhodná pro další zpracování, poté provede korekci na baseline, znormalizuje data podle zvolené metody a v poslední řadě odstraní šum. Výsledkem je analýza dat, která může být obohacena o další knihovny a balíčky s funkcemi přinášející řadu grafů a statistických výsledků. Proto bude tato knihovna a soubor jejich funkcí použit při úpravě dat [29].

### 6.1 Předúprava naměřených dat k využití v dalších analýzách

Na obrázku 6.1 je zobrazeno vybrané spektrum, které bylo získáno proměřením vzorku zaschlé plazmy za pomoci RS za dodržení konstantních podmínek, které jsou uvedené v předchozí kapitole. Vzorek s označením MUZ\_3433\_37 představuje muže s identifikačním číslem 3433 a věkem 37 let. Takto neupravená spektra by bylo nemožné vzájemně porovnávat, proto je nutné přistoupit k několika krokům předúpravy naměřených dat, mezi které patří:

1. vybrání spektrální oblasti,
2. odstranění spektrálního pozadí,
3. dále normalizace pomocí vybraných metod,

## 4. odstranění šumu.



Obrázek 6.1: Naměřené spektrum plazmy s označením MUZ\_3433\_37

## 6.1.1 Vybrání spektrální oblasti

V prvním kroku bylo tedy nutné vybrat na základě charakteru naměřených dat takovou spektrální oblast, která by vykazovala maximum variability mezi vzorky. Již na první pohled bylo zřejmé, že největší variabilita naměřených spekter se objevovala v rozmezí od 700 do 1700  $\text{cm}^{-1}$ . V této škále bylo nalezeno několik pásů, které by pro analýzu mohly být významné viz 6.1. Spektrální oblast nad 1700  $\text{cm}^{-1}$  nebyla vybrána, protože nenesou specifické informace. Ramanova spektra mají obecně pro praktickou klasifikaci příliš mnoho dat, proto se z naměřených dat porovnávají spektra pouze v rozmezí určité frekvence a přebytečné části spekter tak mohou být odstraněny. V této práci se jedná konkrétně o redundantní data z oblasti valenční vibrace CH a hluché místo, kdy k odstranění poslouží funkce `removeFreq`, která je součástí knihovny `chemoSpec`.

## 6.1.2 Odstranění spektrálního pozadí

Dalším potřebným krokem se při zpracování dat ukázalo odstranění spektrálního pozadí za pomoci odečtení polynomu nejnižšího stupně od spekter. U naměřených Ramanových spekter si můžeme povšimnout určitého trendu v rámci jeho pozadí. Při běžných analýzách je snaha naleznout takovou polynomiální funkci, která po odečtení z původního spektra ukáže čisté spektrum bez jakéhokoliv spektrálního pozadí, jakým je nejčastěji fluorescenční pozadí. Řada zkoumaných látek totiž fluoreskuje a přináší rušivý signál, který je nutné pro další kvantitativní analýzu odstranit [30].

V rámci využití spektrální knihovny `chemoSpec` bylo možné odstranit spektrální pozadí za pomoci balíčku `baselineSpectra` obsahující funkci `lowpass` a `modpolyfit`. Z celé řady možností byly vybrány právě tyto dvě vzhledem k nejlepším referencím k eventuálnímu použití. Metoda `lowpass` využívá algoritmu filtrování pomocí Fourierových transformací, v literatuře ji lze najít pod označením Low-pass FFT filter [31]. Funkce `modpolyfit`, jejíž popis lze nalézt [32], pracuje na principu polynomiálního fittování s potlačením základní linie vzhledem k původnímu spektru.

### 6.1.3 Normalizace dat

Nedílnou součástí předzpracování dat je bezesporu proces normalizace. Poslouží jako nástroj snazší porovnatelnosti spekter stejných materiálů mezi sebou. Jednotlivé vzorky, ačkoliv se jedná o ten samý materiál, nemusí být proměřeny za stejných podmínek přístroje, světla, výkonu nebo intenzity laseru a mohou se naměřená data lišit. Normalizace je proces, který vede k zajištění co nejpodobnějších výsledků rozdílných úrovní intenzity, které se objevují i při použití stejného materiálu. Snahou je, aby intenzita daného Ramanova pásu vykazovala co nejpodobnější výsledky napříč všech spekter. Použití různých normalizačních technik vykazuje odlišné výsledky na základě odlišného charakteru naměřených dat [33].

V této práci bude použit balíček normalizačních funkcí `normSpectra`, který je opět součástí knihovny `chemoSpec` nabízející řadu funkcí: `PQN`, `TotInt`, `Range`, `zeroZone`, které byly popsány v článku [34]. Jednotlivé funkce je možné použít pro jiné typy dat, proto bude nutné vyzkoušet, která z výše vypsanych funkcí by byla potencionálně nejlepší pro tuto spektrální analýzu.

1. `PQN` provádí pravděpodobnostní normalizaci kvocientů, což by velice zjednodušeně znamená hledání faktoru podílu zkušebního a referenčního spektra. Tato funkce je popisována jako nejuniverzálnější volba pro mnoho datových souborů.
2. `TotInt` normalizuje podle celkové intenzity. Funkční hodnoty spektra jsou normalizovány součtem všech funkčních hodnot uvažovaného spektra.
3. `Range` umožňuje něco podobného jako `TotInt`, ale namísto použití součtu celého spektra se použije pouze součet daného rozsahu. Použití této funkce by bylo vhodné při použití vnitřního standardu, který by byl bez interferencí s potřebou normalizovat data vůči němu.
4. `zeroZone` není vhodný pro zpracování spektrálních dat.

Z uvedeného výčtu tedy vyplývá, že pro zpracování naměřených spekter budou vhodnými kandidáty na použití pouze funkce `TotInt` a `PQN` [34].

### 6.1.4 Odstranění šumu

Šumem se rozumí rušivý nahodilý signál neobsahující žádnou potřebnou analytickou informaci. Proto je nutné přistoupit k jeho odstranění a zamezení výskytu šumu, k čemuž se často využívá Savitsky-Golayova metoda popsána v [35], která se řadí mezi vyhlazovací metodu vedle transformačních metod. Metoda vyhlazování provádí lokální polynomiální regresi na sérii hodnot, které jsou považovány za stejně vzdálené, a určuje vyhlazenou hodnotu pro každý bod. K dispozici jsou rovněž metody pro výpočet derivací, u kterých je nutno klást důraz na vhodné zvolení stupně derivace, aby nedošlo k odstranění významného Ramanského pásu [33].

V rámci předúpravy konkrétních dat byl využit balíček `sgfSpectra` s použitím různých stupňů derivace (nultá, první, druhá,..), který je součástí knihovny `chemoSpec` [29].

## 6.2 Další zpracování naměřených dat

Poté, co byla jednotlivá spektra předupravena, mohla být spektra dále zpracována za pomoci dalších statistických metod.

### 6.2.1 Analýza hlavních komponent – PCA

Principal Components Analysis (PCA) neboli Analýza hlavních komponent je nedílnou součástí explorační analýzy dat sloužící ke snížení počtu dimenzí mnohorozměrných dat. To může být přínosné z hlediska vizualizace, kvůli úspoře času nebo paměti při strojovém učení. Nepracuje s žádnými předpoklady k rozřazení dat do skupin, ale poskytuje klíčové informace k pochopení dat. Metodu analýzy hlavních komponent je možné si představit jako určení minimálního počtu komponent, které jsou nutné k popisu dat, což v zásadě znamená odstranění šumu a nadbytečných informací, ale zároveň pokrytí maxima informací a variability v datech [36].

### 6.2.2 Analýza hierarchického shlukování – HCA

Clustering neboli shlukování je metoda, která rozřazuje data do skupin podle jejich podobnostních charakteristik, zároveň vypočítává vzdálenosti mezi vzorky a zobrazuje je do grafického znázornění. K dispozici je soubor dat, který obsahuje pouze datové body bez jakýchkoliv štítků tříd. Metoda hierarchického shlukování umožňuje získat možnosti rozdělení dat do jednotlivých shluků. Nejprve pracuje s jedním velkým shlukem, při každé iteraci se shluky podobných dat spojují s jinými, až dojde k počtu  $n$  shluků. Tímto způsobem se tedy data rozdělí. Výstupem je tzv. dendrogramu, který vypadá jako strom zaznamenávající jednotlivá sloučení a rozdělení. Svislá osa představuje číselnou vzdálenost mezi vzorky a vodorovná osa ukazuje rozdělení vzorků od sebe [37].

### 6.2.3 Algoritmus nejmenších parciálních čtverců – PLS

Zkratkou PLS se označuje algoritmus nejmenších parciálních čtverců (Partial least square), který se často využívá v diskriminační analýze, která je známá pod zkratkou PLS-DA (částečná diskriminační analýza metodou nejmenších čtverců). Jedná se o vícerozměrnou statistickou metodu používanou pro klasifikaci dat s cílem předpovědět příslušnost k dané třídě. V posledních letech PLS-DA ukazuje velký úspěch při práci s vysokou dimenzí dat pro různé účely, např. při klasifikaci nemocí v lékařské diagnostice a analýze důkazů ve forenzních vědách [38].

### 6.2.4 ROC grafy

ROC (Receiver Operating Characteristics) je pravděpodobnostní křivka využívaná při snaze klasifikovat sadu vzorků na základě jejich podobnosti. ROC křivka určuje závislost specifity na senzitivitě. AUC (Area Under the Curve) představuje stupeň nebo míru separability. Jak název napovídá, jedná se o plochu pod křivkou, která dosahuje hodnoty v intervalu  $\langle 0, 1 \rangle$  a čím vyšší číslo hodnota dosahuje, tím lépe dokáže vzorky od sebe rozlišit. V případě, že je hodnota rovna 1, můžeme s jistotou říci, že vzorky jde

se 100% pravděpodobností rozdělit do předpokládaných tříd. Pokud je hodnota rovna 0, potom nebyla nalezena žádná podobnost mezi vzorky, tudíž je nemožné vzorky od sebe jakkoliv separovat [39].



## Kapitola 7

# Zhodnocení naměřených dat

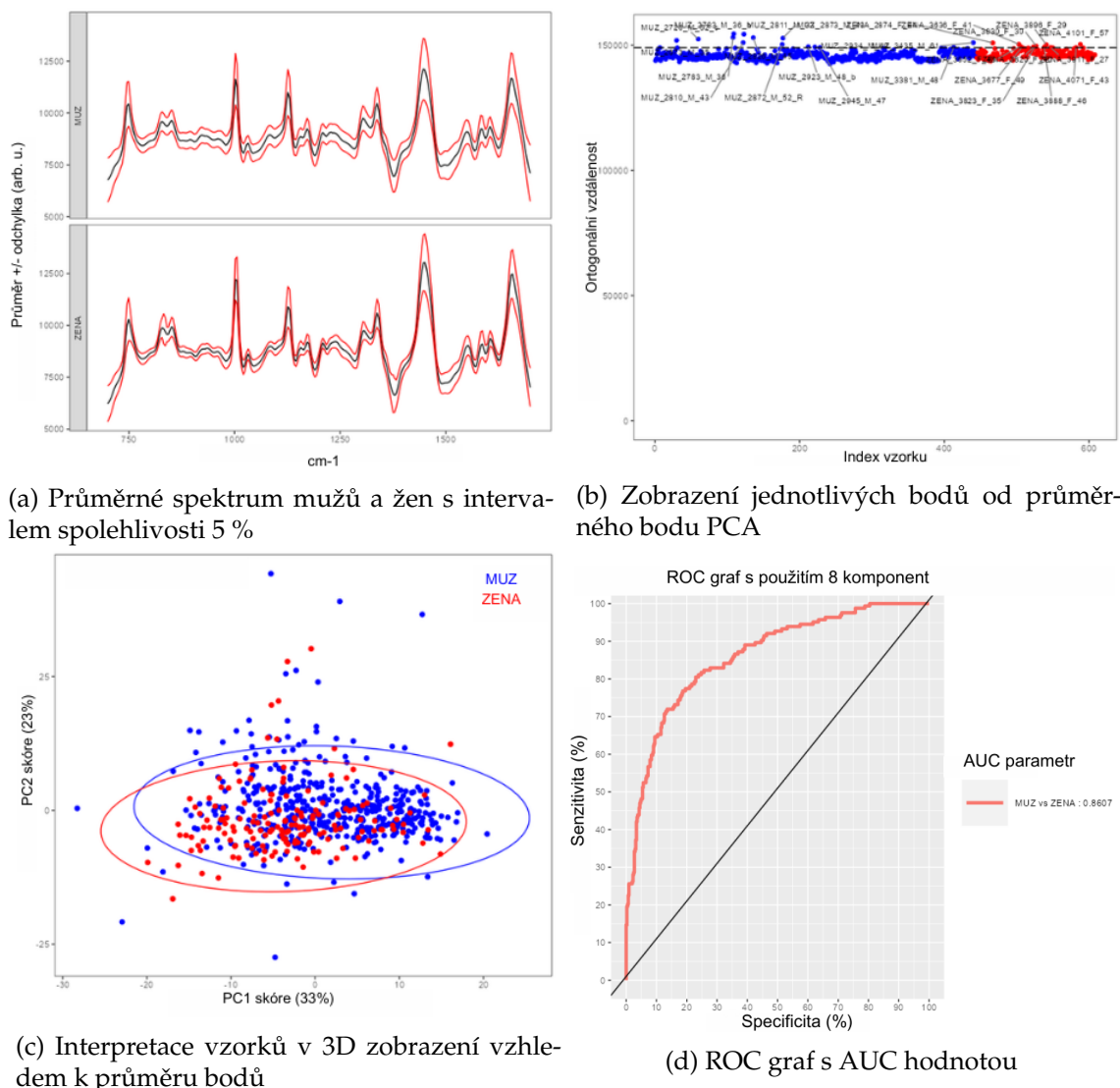
Následující kapitola popisuje získané výsledky, zpracování jednotlivých dat a jejich odlišnosti v důsledku implementovaných funkcí a pozměněných parametrů ve skriptu použitého k analýze dat.

### 7.1 Vybrání vhodné metody pro odstranění spektrálního pozadí

#### 7.1.1 Lowpass

Veškeré použité funkce byly součástí knihovny `chemoSpec`, z níž byla nejprve použita metoda `lowpass` k odstranění spektrálního pozadí ve funkci `baselineSpectra` s normalizační metodou `PQN` ve funkci `normSpectra`. Zároveň bylo použito celkem 8 komponent, které maximálně pokrývají variabilitu všech vzorků. Důvodem použití tohoto počtu komponent, a z jakých výsledků bylo vycházeno bude popsáno na konci této kapitoly. Je také důležité zmínit, že jednotlivé úpravy i další zpracování dat (PCA analýza, HCA analýza) probíhaly současně, ale v této práci budou popisovány odděleně.

Obrázek 7.1a zobrazuje průměrná spektra mužů a žen, přičemž černé pásy reprezentují průměrná Ramanská spektra a červené pásy zobrazují interval spolehlivosti v intervalu  $\pm 5\%$ . Obrázek ortogonální vzdálenosti 7.1b představuje vzdálenost jednotlivých bodů od průměrné hodnoty PCA analýzy, přičemž nejvíce vzdálený vzorek od černé přímky se svým Ramanským spektrem liší nejvíce. Odlehlé vzorky je možné díky jejich označení identifikovat pomocí pohlaví (MUZ/ZENA), čtyřmístného kódu vzorku a věku tvořící poslední dvojmístné číslo v názvu vzorku. Jejich identifikace je nutná pro další zpracování dat. Obrázek s označením 7.1c zobrazuje stejné ukazatele pomocí odlišné interpretace v 3D jako obrázek předchozí, ale bez označení specifických vzorků. Lze zde pozorovat určitý překryv jednotlivých bodů odlišených opět barevnou kombinací modré a červené barvy. Modré body, tj. vzorky mužů vykazují větší rozptyl jednotlivých spekter od sebe, než je tomu u bodů červené barvy patřící ženám. Poslední ze sady obrázků je ROC graf 7.1d ukazující závislost citlivosti na specificitě. Hodnota AUC parametru 0,8607 vyjadřuje zintegrovanou plochu právě pod ROC grafem a ukazuje, s jakou pravděpodobností je možné jednotlivé vzorky od sebe odlišit právě podle pohlaví.

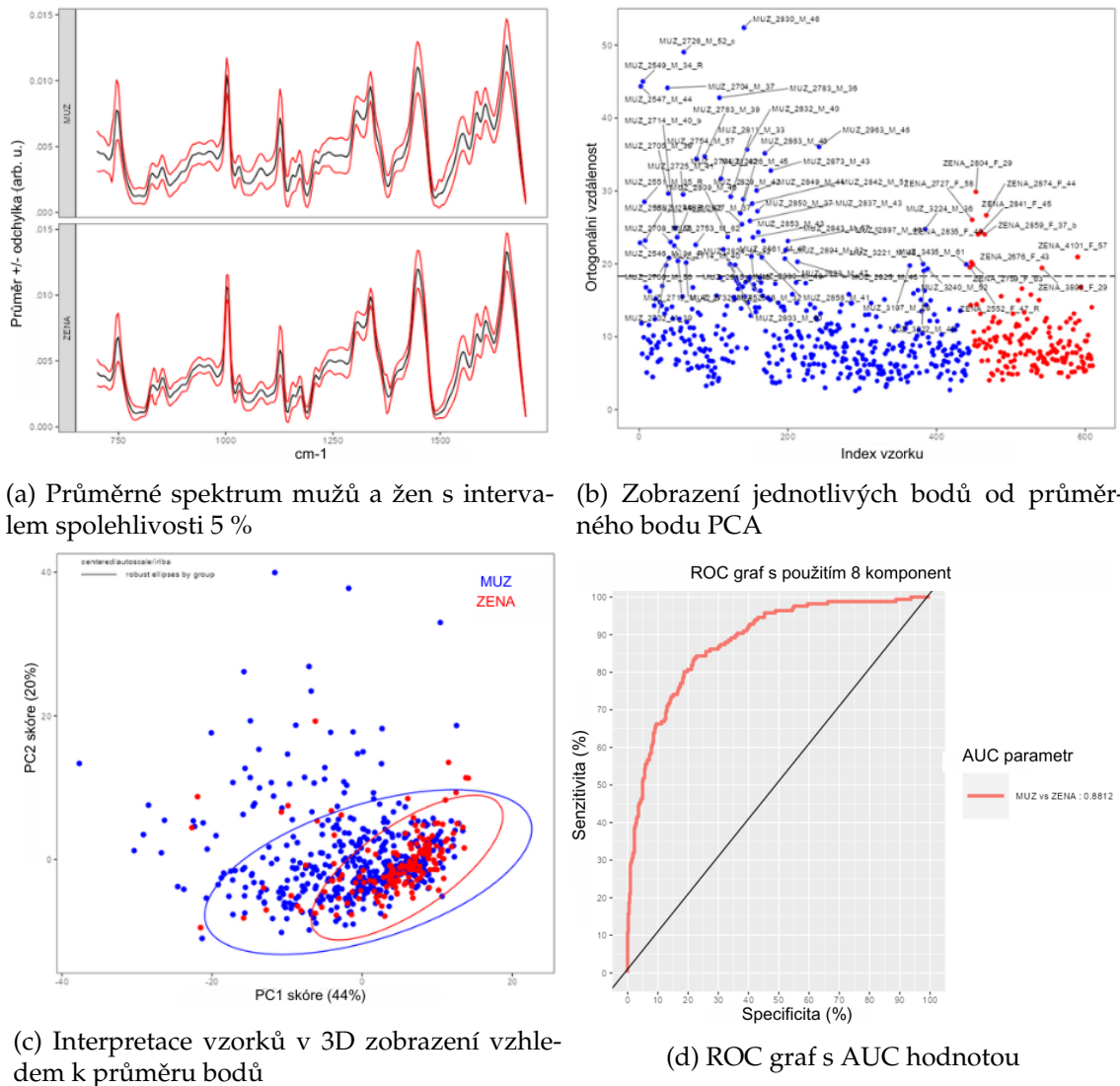


Obrázek 7.1: Zhodnocení naměřených dat pomocí metody lowpass

### 7.1.2 Modpolyfit

Využitá funkce `modpolyfit` zobrazená na obrázcích 7.2, která se často využívá při analýze vibračních technik, na první pohled možná nevykazuje takových změn, ale při bližším pohledu je zlepšení oproti funkci `lowpass` zřejmé. Na obrázku 7.2a jednotlivé pásy vykazují lepší selektivitu, čímž zabraňují většímu překrývání jednotlivých pásů. Ortogonální vzdálenost je v tomto případě lépe pozorovatelná vzhledem ke zvětšenému měřítku na ose y. Bodů, které se od průměrné hodnoty liší je stále značný počet, ale zároveň lze na grafu 7.2c pozorovat lepší překryv spekter. Z výše uvedeného vyplývá, že funkce `modpolyfit` lépe kopíruje fluorescenci, která spektrální pozadí primárně tvoří. V neposlední řadě došlo také ke zvýšení AUC hodnoty na 0,8812. Ačkoliv vykazuje lepší výsledek použití funkce `modpolyfit`, bude i nadále použita funkce `lowpass` při dalších úpravách, aby došlo k potvrzení tohoto úsudku.





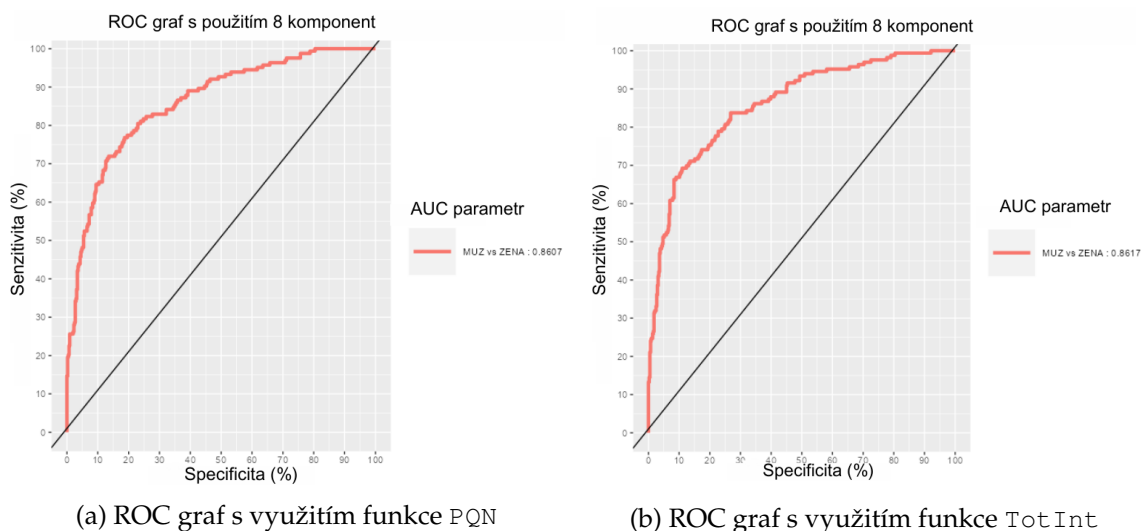
Obrázek 7.2: Zhodnocení naměřených dat pomocí metody modpolyfit

## 7.2 Volba využití funkce pro normalizaci dat

Po výběru vhodné metody pro odstranění spektrálního pozadí bylo nutné přistoupit k otestování normalizačních funkcí, které knihovna `chemoSpec` s funkcí `normSpectra` nabízí. Normalizační funkce `PQN` a `TotInt` byly porovnány v rámci využití s metodami `lowpass` a `modpolyfit`.

### 7.2.1 Využití funkce `PQN` a `TotInt` s použitím metody `lowpass`

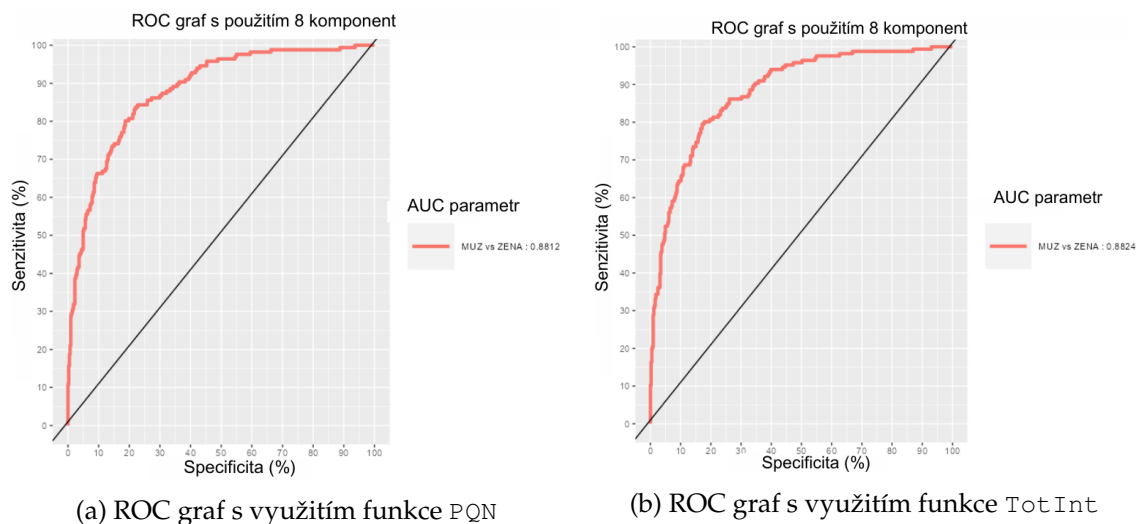
Ačkoliv metoda `lowpass` nevykazovala takové výsledky při volbě ideální funkce pro odstranění spektrálního pozadí, bylo nutné otestovat varianty jejího použití nejen s funkcí `PQN`, ale také s funkcí `TotInt`. Oproti použité funkci `PQN` v minulé podkapitole viz 7.3a s hodnotou 0,8607 došlo v ke zlepšení AUC hodnoty na 0,8617 viz obrázek 7.3b.



Obrázek 7.3: Srovnání ROC grafu a AUC hodnot při použití funkce `lowpass`

## 7.2.2 Využití funkce `PQN` a `TotInt` s použitím metody `modpolyfit`

U metody `PQN`, která byla použita již při odstraňování spektrálního pozadí, byla hodnota AUC (obrázek 7.4a) rovna 0,8812. Při použití normalizační metody `TotInt` změna AUC hodnoty nebyla tak markantní, ale opět došlo k malému zlepšení na 0,8824 viz obrázek 7.4b.



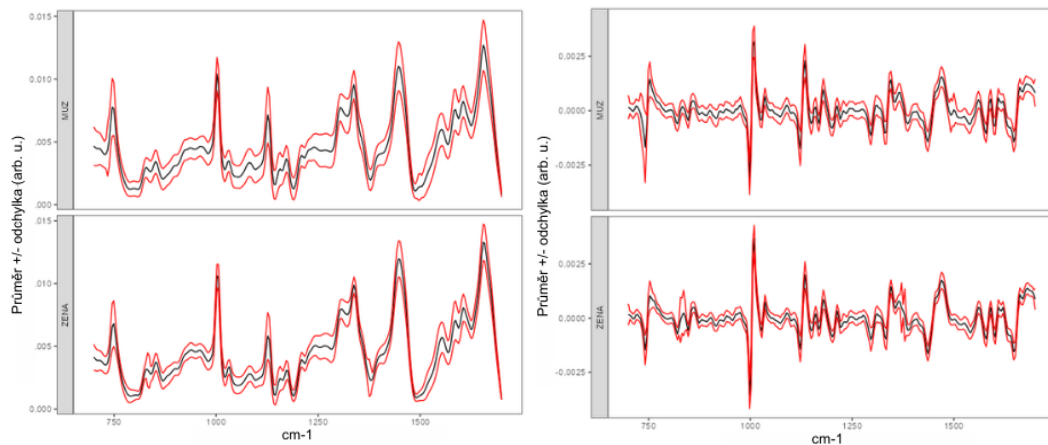
Obrázek 7.4: Srovnání ROC grafu a AUC hodnot při použití funkce `modpolyfit`

Z výše uvedených grafů vyplývá, že normalizační metoda `TotInt` je v kombinaci s oběma funkcemi účinnější než metoda `PQN`. Nejlepších výsledků bylo dosaženo v kombinaci funkce `modpolyfit` a metody `TotInt`.

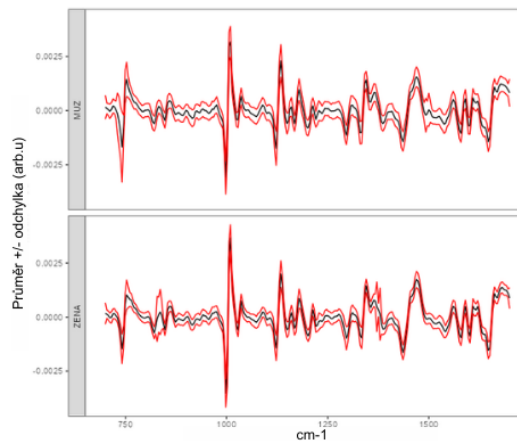
### 7.3 Zvolení vhodného derivačního stupně

V předchozích částech kapitoly byla vybrána vhodná metoda pro zkvalitnění dosažených výsledků. Následně bylo provedeno odstranění šumu za pomoci Savitsky-Golayovy metody za použití funkce `sgfSpectra`, která umožňuje mimo jiné i zvolení vhodného derivačního stupně (první, druhá derivace), ale také využití klouzavého průměru. Obě varianty vedou k lepší projekci dat, což může potenciálně zvýšit kvalitu dosažených výsledků.

Z obrázků průměrných spekter 7.5 vygenerovaných pomocí metody `modpolyfit` a funkcí `TotInt` je zřejmé, že došlo k derivaci křivek, a to pomocí klouzavého průměru (automaticky určuje stupeň derivace), prvního a druhého stupně derivace. Jednotlivá spektra ukazují, že při použití derivačních stupňů nedošlo ke ztrátě žádných pásů, což se při zvolení špatného stupně derivace může stát. Z AUC hodnot zanesených do tabulky 7.1 vyplývá, že použití prvního derivačního stupně nejlépe projektuje naměřená data.



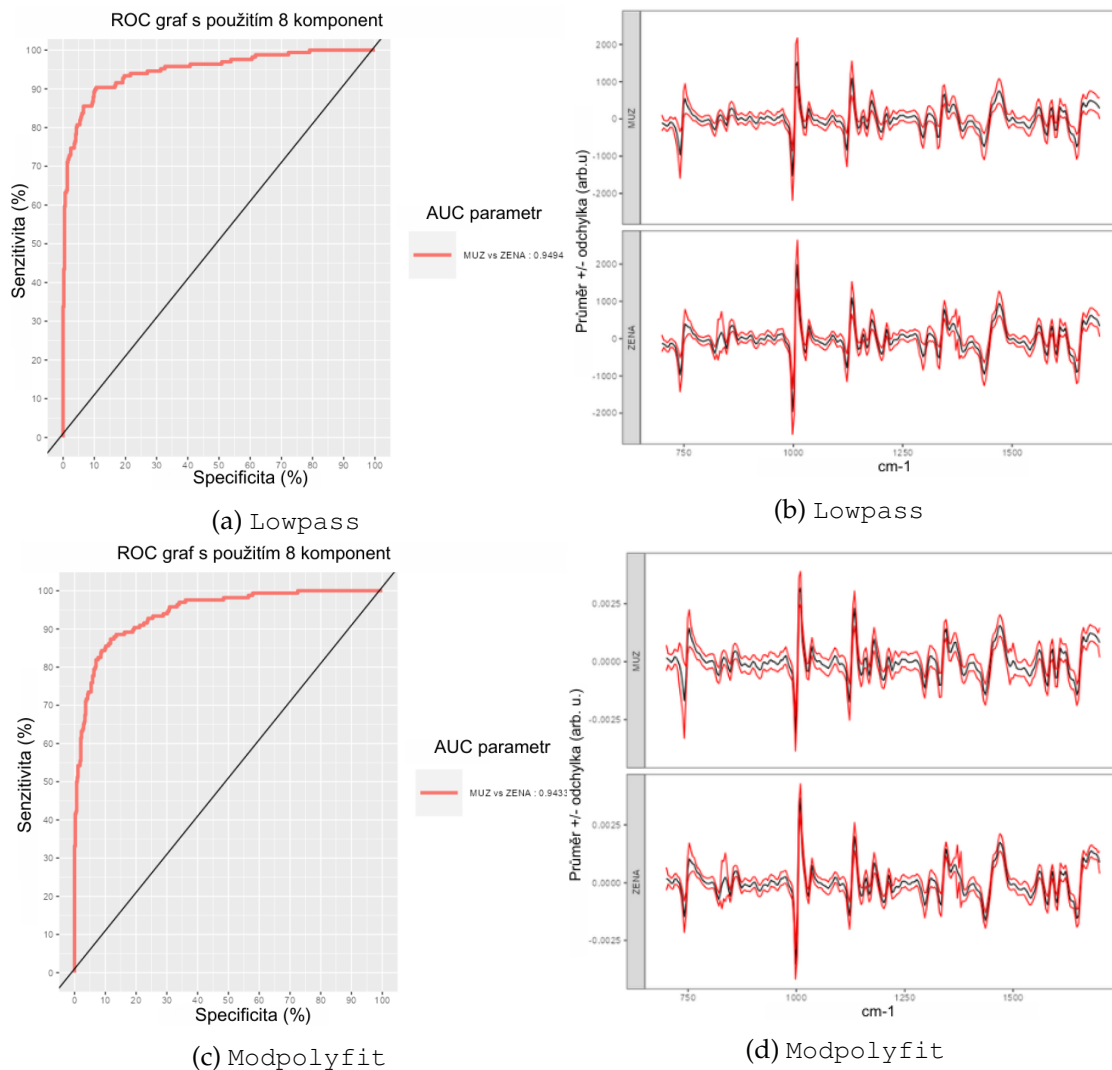
(a) Průměrné spektrum mužů a žen s intervalem spolehlivosti 5 % při použití klouzavého průměru (b) Průměrné spektrum mužů a žen s intervalem spolehlivosti 5 % při použití první derivace



(c) Průměrné spektrum mužů a žen s intervalem spolehlivosti 5 % při použití druhé derivace

Obrázek 7.5: Srovnání vyhlazovací metody při použití klouzavého průměru a při použití první a druhé derivace

Použití prvního derivačního stupně se ukázalo jako nejlepší při použité metodě `modpolyfit`, ale i v kombinaci s metodou `lowpass`, která přinesla nečekaně dobré výsledky. Hodnota AUC byla dokonce vyšší, než při použití metody `modpolyfit`, ačkoliv během dosavadních úprav známky lepších výsledků nevykazovala.



Obrázek 7.6: Srovnání ROC grafů a průměrných spekter s využitím první derivace při metodách `lowpass` a `modpolyfit`

Z dosavadních výsledků lze konstatovat, že oba způsoby vyhodnocení dat poskytují relevantní výsledky. U analýz, které poskytují AUC hodnotu vyšší než 0,7 lze prohlásit, že existuje dostatečně velká pravděpodobnost na správné rozdělení výsledků do požadovaných kategorií. Čím větší je AUC hodnota, tedy plocha pod ROC křivkou, tím lepší selekci získáváme. Nejlepší výsledky této úpravy dat dosahovaly poměrně vysoké AUC hodnoty 0,9433 a 0,9494 zobrazené v tabulce 7.1 [39].

Tabulka 7.1: Souhrnná tabulka s AUC hodnotami po předúpravě dat při použití rozdílných metod pro odstranění spektrálního pozadí a při použití funkcí při normalizaci dat

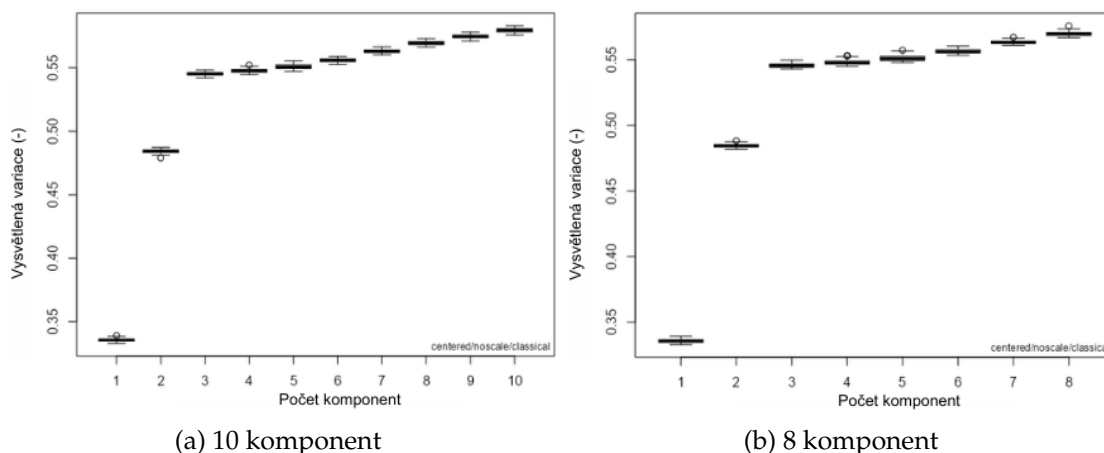
metoda	funkce	stupeň derivace (m)	AUC (%)
<b>lowpass</b>	PQN		86,07
	TotInt		86,17
<b>smooth</b>		0	86,17
		1	<b>94,94</b>
		2	94,31
<b>modpolyfit</b>	PQN		88,12
	TotInt		88,24
<b>smooth</b>		0	88,24
		1	<b>94,33</b>
		2	93,18

## 7.4 Další zpracování dat po jejich předešlé úpravě

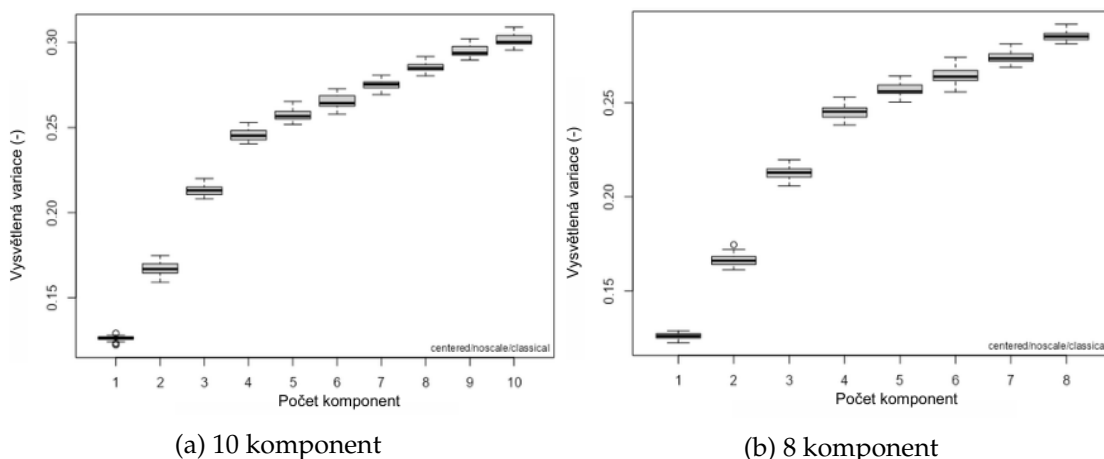
Po procesu předúpravy dat mohlo dojít k dalšímu zpracování naměřených dat za pomoci vybraných analýz (PCA, HCA, clustering). Pro všechny analýzy byla vybrána data s nejlepšími dosaženými výsledky po předúpravě dat, které byly podrobně popsány v předešlé podkapitole.

### 7.4.1 Analýza hlavních komponent – PCA

Při analýze hlavních komponent, která byla do této práce zařazena k nalezení metody pro rozpoznání demografických parametrů, bylo nutné nejprve určit minimální počet komponent, které by vystihly soubor dat co nejlépe.



Obrázek 7.7: Srovnání zvoleného počtu komponent s použitím funkce `lowpass`

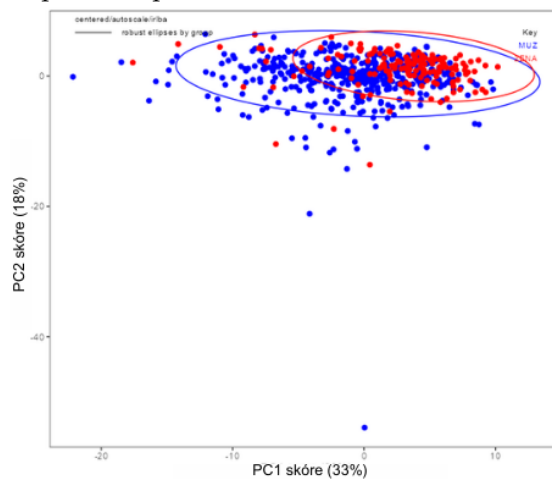
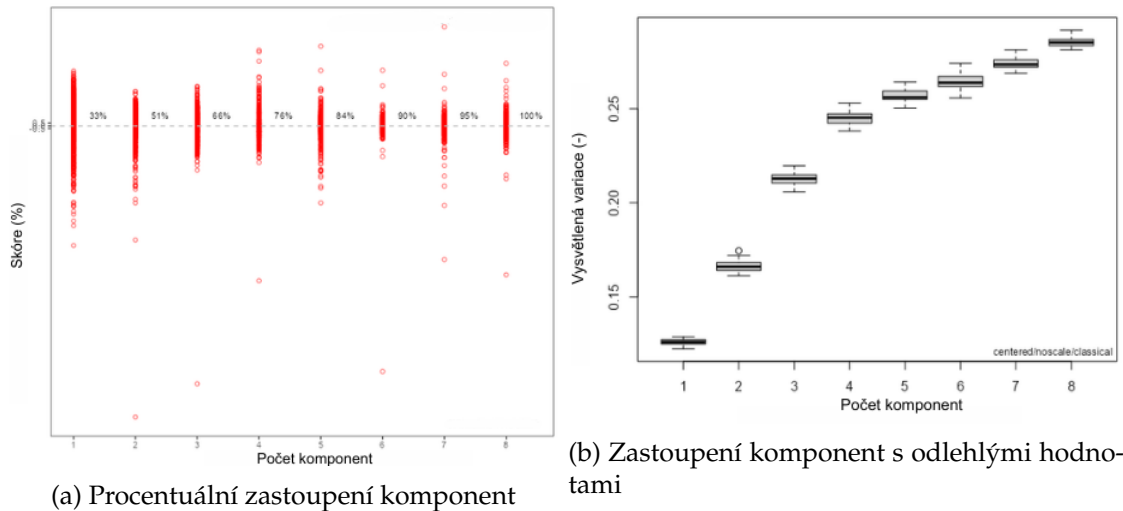


Obrázek 7.8: Srovnání zvoleného počtu komponent při použití funkce `modpolyfit`

Z obrázků 7.7 a 7.8 lze usoudit, že 8 komponent pokrývá co nejvýhodněji celou variabilitu vzorků a není nutné další komponentu přidávat, protože by již zásadně nepomohla. Pomyslná přímka proložená krabicovým grafem po 4. komponentě již zásadně

nevzrůstá. Zároveň je z tohoto obrázku očividné, že metoda `lowpass` nezcela rovnoměrně pokrývá soubor dat oproti metodě `modpolyfit`. I z tohoto důvodu bude dále použita metoda `modpolyfit`.

Výstup PCA analýzy s použitím metody `modpolyfit` zobrazuje série obrázků 7.9, kdy obrázek 7.9a navíc doplňuje předpoklad použití 8 komponent. Rozdíl mezi 7. a 8. komponentou je pouhých 5 %, a při přidání další komponenty by rozdíl ještě klesl, což zároveň ukazuje obrázek 7.9b s krabicovým grafem zobrazující odlehlé hodnoty.

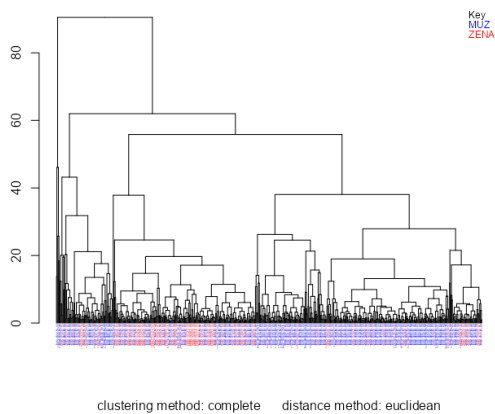


Obrázek 7.9: PCA analýza s použitím metody `modpolyfit`

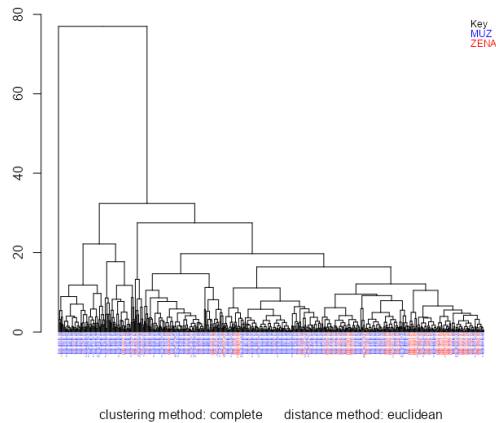
## 7.4.2 Clustering

Obrázky 7.10 ukazují dendrogramy s možnostmi rozdělení datových souborů. V porovnání metod při odstranění spektrálního pozadí `lowpass` zobrazené na obrázku 7.10a a `modpolyfit` (viz 7.10b) si lze povšimnout rozdílu při vytváření jednotlivých červených a modrých oddílů ve spodní části obrázku. V ideálním případě by byly červené vzorky žen uskupeny pohromadě a zároveň by byly odděleny modrými vzorky mužů. Tento trend by byl pozorovatelný pouze v případě, že by bylo možné data rozlišit se 100% pravděpodobností. Touto metodou je možné potvrdit správný výběr při korekci pozadí

spekter tj. `modpolyfit` zobrazuje větší klastery žen, tedy dochází k lepšímu rozdělení dat.



(a) Clustering lowpass



(b) Clustering modpolyfit

Obrázek 7.10: Srovnání klastrování při použití `lowpass` metody s využitím klouzavého průměru a `modpolyfit` s vyhlazením první derivací



## Kapitola 8

# Diskuze a zhodnocení dosažených výsledků

V předchozí kapitole byly zhodnoceny jednotlivé přístupy při předúpravě dat, které byly pro tuto práci stěžejní. Bez jednotlivých kroků by byla data v takovém množství jen těžko porovnatelná a nebylo by možné dosáhnout dostatečně dobrých výsledků. Jednotlivé úpravy byly nutné hlavně z důvodu velkého počtu vzorků, kterých bylo téměř 700. Je také nutné konstatovat, že kvalita spekter byla sama o sobě dost vysoká, což bylo pozorovatelné při prvním náhledu do jednotlivých spekter viz 6.1. Nejlepších výsledků bylo dosaženo při použití metody `lowpass` s funkcí `TotInt` a za pomoci derivace spektra prvního stupně a při využití metody `modpolyfit` s funkcí `TotInt` a rovněž první derivací. Obě varianty úprav spekter by byly adekvátní k dalšímu využití. Ačkoliv byly výsledky při předúpravě dat s použitím funkce `lowpass` nepatrně lepší (především podle AUC paramteru uvedeného v tabulce v předchozí kapitole), při dalších dvou analýzách – analýze hlavních komponent a analýze hierarchického klastrování – se ukázala metoda `modpolyfit` jako vhodnější. Hlavním důvodem k vybrání této metody bylo více rovnoměrné pokrytí souboru dat při použití 8 komponent v PCA analýze viz 7.8 nebo viditelně lepší klastrování zobrazené na dendrogramu viz 7.10b. Za pomoci všech uvedených analýz došlo k rozdělení naměřených spekter klinických vzorků plazmy pomocí Ramanovy spektroskopie na základě vybraného demografického parametru – pohlaví.

Ačkoliv součástí informací o pacientech byl také jejich věk, z důvodu časové náročnosti a rozsahu práce se výzkum tímto parametrem dále nezabýval. I když by bylo nanejvýš zajímavé, jak tento demografický parametr vzorky ovlivňuje, bylo by těžké stanovit jednotlivé věkové hranice, které by měly určovat přechod mezi nimi. Toto téma by mohlo být předmětem dalšího výzkumu v oblasti Ramanovy spektroskopie.



# Závěr

Nalézt vhodný algoritmus, který by dokázal rozpoznat proměřené kapky plazmy Ramanovou spektroskopií (RS) podle pohlaví, nebylo tak snadné, jak se na první pohled může zdát. Celkem 689 dárců plazmy tvořilo soubor dat, který byl nutný rozdělit podle demografického parametru – pohlaví. Vzhledem k tak velkému počtu dat byla nejprve snaha zjednodušit a urychlit nanášení vzorků za pomoci přístroje Labcyte Echo 550 Liquid handler. Jeho použití se ale neukázalo jako přínosné, protože při nanášení vzorků rhodaminu 6G, který byl pro první pokusy zvolen, vykazoval chybovost 13 % při nanášení vzorků. Docházelo k nekontrolovanému smývání a posunu kapek, což by při použití reálných vzorků zabraňovalo k jejich identifikaci. Z tohoto důvodu byl přístup nanášení vzorků zamítnut a bylo nutné veškeré vzorky nanést ručně za pomoci automatické pipety.

Po proměření všech vzorků plazmy došlo k vizuální kontrole naměřených spekter, kdy největší variabilita spektrálních pásů se objevovala v oblasti  $700 - 1700 \text{ cm}^{-1}$ , která byla vybrána pro další analýzu. Veškeré úpravy a následné analýzy byly prováděny v programovacím jazyce R. Vzhledem k velkému množství dat bylo nutné se zaměřit na porovnatelnost spekter mezi sebou, k čemuž posloužily funkce z knihovny `chemoSpec`, která je souborem funkcí sloužící pro analýzu spektrálních dat. Z knihovny byly použity metody pro odstranění spektrálního pozadí, funkce pro normalizaci dat a funkce pro vyhlazení dat. Při použití metody `lowpass` pro odstranění spektrálního pozadí fungující na principu Fourierových transformací a funkce `TotInt`, která normalizuje data podle jejich celkové intenzity, vykazovaly výstupy předúpravy dat lepší výsledky. Parametrem, který byl klíčový pro porovnání jednotlivých přístupů, se stala AUC hodnota ukazující zintegrovanou plochu pod křivkou. Tato hodnota ukazuje s jakou pravděpodobností bylo možné data rozdělit. Při prvních výsledcích byla hodnota AUC rovna 86,07 %.

Součástí této práce byly také analýzy umožňující rozřazení dat. Konkrétně se jednalo o PCA (analýzu hlavních komponent), HCA (analýzu hierarchického klastrování) a PLS (algoritmus nejmenších parciálních čtverců). Při PCA analýze bylo nutné určit dostatečný počet komponent, které by pokryly variabilitu naměřených dat. V tomto případě bylo zvoleno 8 komponent, které postačily k maximálnímu pokrytí. Při analýze HCA byly sledovány klastry, které se při hierarchickém klastrování tvořily, což opět napovídalo k možnému rozdělení dat. Za pomoci PLS algoritmů byly generovány ROC grafy, a po zintegrování plochy pod touto křivkou udávaly již zmiňovaný AUC parametr. Po provedených analýzách za pomoci vhodnějších úprav byla hodnota AUC posunuta až k hranici 95 %, konkrétně 95,94 %. S touto pravděpodobností bylo možné rozdělit soubor dat podle pohlaví. V budoucnu by tento druh analýzy mohl sloužit pro forenzní účely a kriminalistiku. Vyvinutý algoritmus by mohl být vhodný pro použití i dalších klinických vzorků za stejným účelem rozdělení vzorků, nebo k analýze jiných demografických parametrů.



# Summary

Finding a suitable algorithm that could recognize Raman spectroscopy (RS) measured plasma droplets by gender was not as easy as it may seem. The dataset comprised of a total of 689 plasma donors, who needed to be divided by a demographic parameter – gender. Because of such a large number of data, an effort was first made to simplify and speed up the sample application using the Labcyte Echo 550 Liquid handler. However, its use did not prove to be beneficial, because the 6G rhodamine sample application device chosen for the first experiments had a 13 % error of application. There was uncontrolled mixing and shifting of droplets, which would have prevented the identification of real samples. For this reason, application of samples approach was rejected and all samples had to be deposited manually using an automated pipette.

After all plasma samples were measured, the measured spectra were visually inspected, with the greatest variability in spectral bands appearing in the  $700 - 1700 \text{ cm}^{-1}$  region, which was selected for further analysis. All adjustments and subsequent analyses were performed in the programming language R. Due to the large amount of data, it was necessary to focus on the comparability of the spectra with each other, which was done by using functions from the library `chemoSpec`, which is a set of functions used for spectral data analysis. From the library, methods for spectral background removal, functions for data normalization and functions for data smoothing were used. When using the `lowpass` method for spectral background removal, which operates on the principle of Fourier transforms, and the `TotInt` function, which normalizes the data according to their total intensity, the data preprocessing outputs showed better results. The parameter that was crucial for comparing the different approaches was the AUC value indicating the integrated area under the curve. This value indicates with what probability the data could be split. In the first results, the AUC value was equal to 86,07 %.

This work also included analyses to allow the data to be sorted. Specifically, these were PCA (principal component analysis), HCA (hierarchical clustering analysis) and PLS (partial least squares). In the PCA analysis, it was necessary to identify a number of components to cover the variability of the measured data. In this case, 8 components were chosen to provide sufficient for maximum coverage. In the HCA analysis, the clusters that formed during hierarchical clustering were observed, which again suggested a possible distribution of the data. ROC plots were generated using PLS algorithms, and when integrated, the areas under this curve gave the aforementioned AUC parameter. After analyses using more appropriate adjustments, the AUC value was moved to the 95 % limit, namely 95,94 %. With this probability, it was possible to split the data set by sex. In the future, this kind of analysis could be used for forensic purposes and criminalistics. The algorithm developed could be suitable for use with other clinical samples for the same purpose of sample distribution, or for analysis of other demographic parameters.



# Bibliografie

1. SMITH, W. E.; DENT, G. *Modern Raman Spectroscopy – A Practical Approach*. John Wiley & Sons, 2005. ISBN 0-471-49668-5 (HB); 0-471-49794-0 (PB).
2. PETERSEN, M.; YU, Z.; LU, X. Application of Raman Spectroscopic Methods in Food Safety: A Review. *Biosensors* [online]. 2021, vol. 11, no. 6, s. 187 [cit. 2023-01-18]. ISSN 2079-6374. Dostupné z DOI: 10.3390/bios11060187. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.
3. METAXIA, P. et al. ARCHAEOLOGICAL AND HISTORICAL STUDY OF LYKION COMPLEMENTED BY IR AND RAMAN SPECTROSCOPIC INVESTIGATION. *Scientific Culture* [online]. 2022, roč. 8, č. 1, s. 95–114 [cit. 2023-01-18]. ISSN 24080071. Dostupné z: <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=24080071&v=2.1&it=r&id=GALE%7CA697851486&sid=googleScholar&linkaccess=abs>. Publisher: University of the Aegean.
4. DOTY, K. C. et al. What can Raman spectroscopy do for criminalistics? *Journal of Raman Spectroscopy* [online]. 2016, vol. 47, no. 1, s. 39–50 [cit. 2023-01-18]. ISSN 1097-4555. Dostupné z DOI: 10.1002/jrs.4826.
5. SHIPP, D. W.; SINJAB, F.; NOTINGHER, I. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics* [online]. 2017, roč. 9, č. 2, s. 315–428 [cit. 2021-11-29]. ISSN 1943-8206. Dostupné z DOI: 10.1364/AOP.9.000315. Publisher: Optical Society of America.
6. WIDJAJA, E.; HONG LIM, G.; AN, A. A novel method for human gender classification using Raman spectroscopy of fingernail clippings. *Analyst* [online]. 2008, vol. 133, no. 4, s. 493–498 [cit. 2023-04-03]. Dostupné z DOI: 10.1039/B712389B. Publisher: Royal Society of Chemistry.
7. GAMULIN, O. et al. Possibility of Human Gender Recognition Using Raman Spectra of Teeth. *Molecules* [online]. 2021, roč. 26, č. 13, s. 3983 [cit. 2023-01-16]. ISSN 1420-3049. Dostupné z DOI: 10.3390/molecules26133983.
8. WOLTHUIS, R. et al. Raman Spectroscopic Methods for In Vitro and In Vivo Tissue Characterization. In: *Fluorescent and Luminescent Probes for Biological Activity* [online]. Elsevier, 1999, s. 433–455 [cit. 2021-11-28]. ISBN 978-0-12-447836-7. Dostupné z DOI: 10.1016/B978-012447836-7/50034-8.
9. ROSTRON, P.; GERBER, D. Raman Spectroscopy, a review. *International Journal of Engineering and Technical Research*. 2016, roč. 6, s. 50–64.
10. SMEKAL, A. Zur Quantentheorie der Dispersion. *Naturwissenschaften* [online]. 1923, roč. 11, s. 873–875 [cit. 2023-04-04]. Dostupné z DOI: <https://doi.org/10.1007/BF01576902>.

11. KHAN, A. Science Reporter: The Phenomen of Raman. 1988, vol. Vol 25, s. 38. Dostupné také z: <http://dspace.rri.res.in/bitstream/2289/7704/3/Science%20Reporter-%20Nov-%20Dec%201988.pdf>.
12. BAKER, M. *Biophotonics: Vibrational Spectroscopic Diagnostics* [online]. IOP Publishing, 2016 [cit. 2021-11-28]. ISBN 978-1-68174-071-3. Dostupné z DOI: 10.1088/978-1-6817-4071-3.
13. FERRARO, J. R.; NAKAMOTO, K.; BROWN, C. W. Chapter 1 - Basic Theory. In: FERRARO, J. R.; NAKAMOTO, K.; BROWN, C. W. (eds.). *Introductory Raman Spectroscopy (Second Edition)* [online]. San Diego: Academic Press, 2003, s. 1–94 [cit. 2022-11-21]. ISBN 978-0-12-254105-6. Dostupné z DOI: 10.1016/B978-012254105-6/50004-4.
14. SMITH, G. D.; CLARK, R. J. H. Raman microscopy in archaeological science. *Journal of Archaeological Science* [online]. 2004, vol. 31, no. 8, s. 1137–1160 [cit. 2023-01-19]. ISSN 0305-4403. Dostupné z DOI: 10.1016/j.jas.2004.02.008.
15. PENCE, I.; MAHADEVAN-JANSEN, A. Clinical instrumentation and applications of Raman spectroscopy. *Chemical Society Reviews* [online]. 2016, vol. 45, no. 7, s. 1958–1979 [cit. 2023-01-19]. Dostupné z DOI: 10.1039/C5CS00581G. Publisher: Royal Society of Chemistry.
16. ELUMALAI, S.; MANAGÓ, S.; DE LUCA, A. C. Raman Microscopy: Progress in Research on Cancer Cell Sensing. *Sensors* [online]. 2020, vol. 20, no. 19, s. 5525 [cit. 2023-01-19]. ISSN 1424-8220. Dostupné z DOI: 10.3390/s20195525. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.
17. *Jedovatý alkohol zabil tři lidi, otrávených přibývá - Deník.cz* [online]. 2012. [cit. 2023-01-18]. Dostupné z: [https://www.denik.cz/z\\_domova/jedovaty-alkohol-zabil-tri-lidi-otravenych-pribyva-20120910.html](https://www.denik.cz/z_domova/jedovaty-alkohol-zabil-tri-lidi-otravenych-pribyva-20120910.html).
18. ZIEMANN, M. A.; MADARIAGA, J. M. Applications of Raman spectroscopy in art and archaeology. *Journal of Raman Spectroscopy* [online]. 2021, roč. 52, č. 1, s. 8–14 [cit. 2023-01-18]. ISSN 1097-4555. Dostupné z DOI: 10.1002/jrs.6054. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.6054](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.6054).
19. AL-HETLANI, E. et al. Differentiating smokers and nonsmokers based on Raman spectroscopy of oral fluid and advanced statistics for forensic applications. *Journal of Biophotonics* [online]. 2020, vol. 13, no. 3, e201960123 [cit. 2023-01-18]. ISSN 1864-0648. Dostupné z DOI: 10.1002/jbio.201960123.
20. STARITZBICHLER, R. et al. Raman spectroscopy on blood serum samples of patients with end-stage liver disease. *PLoS ONE* [online]. 2021, roč. 16, č. 9, e0256045 [cit. 2023-01-18]. ISSN 1932-6203. Dostupné z DOI: 10.1371/journal.pone.0256045.
21. LIMA, A. et al. Discrimination of leukemias and non-leukemic cancers in blood serum samples of children and adolescents using a Raman spectral model. *Lasers in Medical Science*. 2023, roč. 38, č. 1. ISSN 0268-8921. Dostupné z DOI: 10.1007/s10103-022-03681-2.
22. GUALERZI, A. et al. Raman profiling of circulating extracellular vesicles for the stratification of Parkinson's patients. *Nanomedicine: Nanotechnology, Biology and Medicine* [online]. 2019, vol. 22, s. 102097 [cit. 2023-02-16]. ISSN 1549-9634. Dostupné z DOI: 10.1016/j.nano.2019.102097.



23. JADHAV, S. A. et al. Development of integrated microfluidic platform coupled with Surface-enhanced Raman Spectroscopy for diagnosis of COVID-19. *Medical Hypotheses* [online]. 2021, vol. 146, s. 110356 [cit. 2023-03-07]. ISSN 0306-9877. Dostupné z DOI: 10.1016/j.mehy.2020.110356.
24. KOSTER, H. J. et al. Fused Raman spectroscopic analysis of blood and saliva delivers high accuracy for head and neck cancer diagnostics. *Scientific Reports* [online]. 2022, vol. 12, no. 1, s. 18464 [cit. 2023-02-13]. ISSN 2045-2322. Dostupné z DOI: 10.1038/s41598-022-22197-x.
25. RAMAKRISHNAIAH, R. et al. Applications of Raman Spectroscopy in Dentistry: Analysis of Tooth Structure. *Applied Spectroscopy Reviews* [online]. 2015, roč. 50, č. 4, s. 332–350 [cit. 2023-04-03]. ISSN 0570-4928. Dostupné z DOI: 10.1080/05704928.2014.986734.
26. PARACHALIL, D. R.; MCINTYRE, J.; BYRNE, H. J. Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: recent advances. *Analytical and Bioanalytical Chemistry* [online]. 2020, vol. 412, no. 9, s. 1993–2007 [cit. 2023-03-08]. ISSN 1618-2650. Dostupné z DOI: 10.1007/s00216-019-02349-1.
27. KERR, L. T.; BYRNE, H. J.; HENNELLY, B. M. Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. *Analytical Methods* [online]. 2015, vol. 7, no. 12, s. 5041–5052 [cit. 2023-03-08]. ISSN 1759-9679. Dostupné z DOI: 10.1039/C5AY00327J. Publisher: The Royal Society of Chemistry.
28. ZEHENTBAUER, F. M. et al. Fluorescence spectroscopy of Rhodamine 6G: Concentration and solvent effects. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* [online]. 2014, vol. 121, s. 147–151 [cit. 2023-02-16]. ISSN 1386-1425. Dostupné z DOI: 10.1016/j.saa.2013.10.062.
29. HANSON, B. A. *ChemoSpec: Exploratory Chemometrics for Spectroscopy. R package version 4.3.34*. 2016. Dostupné také z: [github.com/bryanhanson/ChemoSpec](https://github.com/bryanhanson/ChemoSpec).
30. CADUSCH, P. J. et al. Improved methods for fluorescence background subtraction from Raman spectra. *Journal of Raman Spectroscopy* [online]. 2013, vol. 44, no. 11, s. 1587–1595 [cit. 2023-02-22]. ISSN 1097-4555. Dostupné z DOI: 10.1002/jrs.4371. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.4371](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.4371).
31. ATAKAN, A. K.; BLASS, W. E.; JENNINGS, D. E. Elimination of Baseline Variations from a Recorded Spectrum by Ultra-low Frequency Filtering. *Applied Spectroscopy* [online]. 1980, roč. 34, č. 3, s. 369–372 [cit. 2023-02-15]. Dostupné z: <https://opg.optica.org/as/abstract.cfm?uri=as-34-3-369>. Publisher: Society for Applied Spectroscopy.
32. LIEBER, C. A.; MAHADEVAN-JANSEN, A. Automated method for subtraction of fluorescence from biological Raman spectra - PubMed [online]. 2003 [cit. 2023-02-15]. Dostupné z DOI: 10.1366/000370203322554518.
33. GAUTAM, R. et al. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation* [online]. 2015, vol. 2, no. 1, s. 1–38 [cit. 2023-02-23]. ISSN 2195-7045. Dostupné z DOI: 10.1140/epjti/s40485-015-0018-6. Number: 1 Publisher: SpringerOpen.
34. DIETERLE, F. et al. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabonomics. *Analytical Chemistry*. 2006, roč. 78, č. 13, s. 4281–4290. ISSN 0003-2700. Dostupné z DOI: 10.1021/ac051632c.

35. PRESS, W. H. (ed.). *FORTRAN numerical recipes*. 2nd ed. Cambridge [England] ; New York: Cambridge University Press, 1996. ISBN 978-0-521-43064-7 978-0-521-57439-6.
36. TASKESEN, E. *What are PCA loadings and how to effectively use Biplots?* [Medium] [online]. 2023-02-22. [cit. 2023-04-18]. Dostupné z: <https://towardsdatascience.com/what-are-pca-loadings-and-biplots-9a7897f2e559>.
37. PATLOLLA, C. R. *Understanding the concept of Hierarchical clustering Technique* [Medium] [online]. 2020-05-29. [cit. 2023-03-06]. Dostupné z: <https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec>.
38. LEE, L. C.; LIONG, C.-Y.; JEMAIN, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* [online]. 2018, vol. 143, no. 15, s. 3526–3539 [cit. 2023-04-03]. ISSN 1364-5528. Dostupné z DOI: 10.1039/C8AN00599K. Publisher: The Royal Society of Chemistry.
39. NARKHEDE, S. *Understanding AUC - ROC Curve* [Medium] [online]. 2021-06-15. [cit. 2023-03-02]. Dostupné z: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.