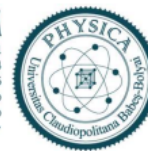




UNIVERSITATEA BABEȘ-BOLYAI
BABEȘ-BOLYAI TUDOMÁNYEGYETEM
BABEȘ-BOLYAI UNIVERSITÄT
BABEȘ-BOLYAI UNIVERSITY
TRADITIO ET EXCELLENTIA

FACULTATEA DE FIZICĂ
Str. Mihail Kogălniceanu nr.1
Cluj-Napoca, RO-400084
Tel: +4(0)264-405300 | FAX: +4(0)264-591906
secretariat.phys@ubbcluj.ro
www.phys.ubbcluj.ro



UNIVERSITATEA “BABEȘ-BOLYAI” CLUJ-NAPOCA

FACULTATEA DE FIZICĂ

SPECIALIZAREA BIOFIZICĂ ȘI FIZICĂ MEDICALĂ

LUCRARE DE DISERTAȚIE

Detecția genei PML-RAR α din ADN prin spectroscopie SERS

Coordonatori științifici

Prof. dr. Nicolae Leopold

Lect. dr. Ștefania Iancu

Absolvent

Havrici Ana Maria

2024

Abstract

The activation of the PML-RAR α gene in DNA affects the response to chemotherapy in acute promyelocytic leukemia, leading to different treatment approaches for patients. However, due to the high cost of identifying this gene through next-generation sequencing, many countries do not treat patients differently based on the presence or absence of this gene. This thesis aims to determine if surface-enhanced Raman spectroscopy (SERS), a fast and inexpensive method, can identify the PML-RAR α gene in DNA samples extracted from patients' blood.

The thesis is divided into four chapters. The first chapter covers theoretical concepts related to DNA, surface-enhanced Raman spectroscopy, and acute promyelocytic leukemia. The second chapter describes the materials and methods used in the study, including DNA samples extracted from blood samples of patients both positive and negative for the PML-RAR α gene. Ag nanoparticles stabilized with hydroxylamine hydrochloride and activated by Ca²⁺ were used as the SERS enhancement substrate. SERS spectra were acquired and processed using Quasar software.

The third chapter presents the results obtained from processing the SERS spectra, examining how DNA samples are influenced by the presence of Ca²⁺, and the effects of freezing and storage on the samples. It also discusses the differences observed between fresh and frozen DNA samples and the development of classification models for these samples. The study found that DNA can be detected by SERS only through its adsorption on silver nanoparticles using Ca²⁺. Moreover, DNA detection by SERS is affected by the storage time of both the blood samples before DNA extraction and the extracted DNA. Finally, the thesis demonstrates that SERS can distinguish between DNA with positive and negative PML-RAR α gene presence with a classification accuracy of 79.6%.

Cuprins

Introducere	5
Capitolul I. Noțiuni teoretice	7
1.1 ADN.....	7
1.2 SERS (Spectroscopia Raman amplificată de suprafață)	9
1.3 Leucemie acută promielocitară.....	13
Capitolul II. Materiale și metode	16
2.1. Probele de ADN.....	16
2.2. Extragerea ADN-ului	16
2.3 Sinteza de nanoparticule de argint	17
2.4 Achiziționarea spectrelor SERS ale probelor de ADN	18
2.5 Procesarea spectrelor	19
2.6 Modele de clasificare	19
Capitolul III. Rezultate	21
3.1. Sinteza AgNPs.....	21
3.2. Detecția SERS a ADN-ului	22
3.2.1. Rolul Ca^{2+} în detecția SERS a ADN-ului.....	22
3.2.2. Influența stocării probelor în achiziționarea spectrelor SERS a ADN-ului	25
3.3. Clasificarea pacienților cu genă PML-RARA pozitivă și negativă pe baza spectrelor SERS ale ADN-ului.....	28
Concluzii	39

Bibliografie40

Introducere

Leucemia promielocitară acută (LPA) este un tip de cancer care se caracterizează creșterea rapidă a celulelor în măduva osoasă și sânge. LPA prezintă o translocăție cromozomială la cromozomii 15 și 17, sau la genele PML și RAR α rezultând o genă de fuziune numită PML-RAR α , care poate fi negativă sau pozitivă.

Detecția genei PML-RAR α se face prin secvențierea ADN-ului, tehnică costisitoare și cu o durată de timp mare pentru identificarea genei, ceea ce limitează accesul la un diagnostic rapid. Tratamentul pentru gena PML-RAR α variază în funcție de tip genei, pozitiv sau negativ. Pacienții diagnosticați cu LPA cu gena PML-RAR α pozitivă urmează un tratament alcătuit din acid all-trans retinoic și arsenic trioxid, îmbunătățind tratamentul pentru această genă și crescând semnificativ rata de supraviețuire. În schimb, pacienții diagnosticați cu LPA cu gena PML-RAR α negativă urmează un tratament mai toxic, cum ar fi chimioterapia.

Dezvoltarea celor mai recente tratamente necesită o infrastructură medicală avansată și acces la o metodă de diagnostic rapidă și precisă, ceea ce poate dura un timp mare de dezvoltare, de aceea în unele țări se utilizează metoda actuală de diagnostic și tratament.

Pentru a putea diagnostica gena PML-RAR α într-un timp mai scurt, trebuie dezvoltate și studiate diferite tehnici care ar putea ajuta la un diagnostic mai rapid. Una dintre tehnicile care ar facilita diagnosticul rapid al genei PML-RAR α este spectroscopia Raman amplificată la suprafață (SERS). SERS este o tehnică care oferă rezultate într-un timp scurt și cu o sensibilitate ridicată, care poate fi utilizată pentru screening-ul biofluidelor și a probelor biofluide extrase din țesut sau sânge.

În lucrarea de față am încercat posibilitatea detecției genei PML-RAR α din probele de ADN cu gena PML-RAR α extrase din sângele pacienților prin metoda SERS și dezvoltarea de modele de clasificare pentru probele de ADN cu gena PML-RAR α .

Lucrarea este alcătuită din patru capitole. Primul capitol acoperă aspectele teoretice referitoare la ADN, spectroscopia Raman amplificată la suprafață și leucemia promielocitară acută.

În al doilea capitol sunt prezentate materialele și metodele utilizate pentru această lucrare. Sunt prezentate probele de ADN utilizate, modalitatea de extragere a ADN-ului din sângele pacienților cu gena PML-RAR α , nanoparticulele de Ag utilizate și sinteza lor, procesul de

achiziționarea spectrelor SERS și spectrometrul utilizat, pașii urmați pentru procesarea spectrelor SERS obținute și o descriere a modelelor de clasificare pe care am dorit să le dezvoltăm.

În al treilea capitol sunt prezentate rezultatele obținute, descrierea și prezentarea nanoparticulelor de Ag, inclusiv spectrul UV-VIS, metoda de detecție a ADN-ului prin SERS. În capitolul este descrisă influența utilizării ionilor de Ca pentru detecția ADN-ului prin metoda SERS și cum interacționează cu nanoparticulele de Ag. Un alt subiect discutat în acest capitol este influența stocării probelor de ADN extras din sânge, influența care se poate observa în semnalul SERS. Capitolul trei se încheie cu clasificarea pacienților cu gena PML-RAR α pe baza spectrelor SERS, clasificare realizată în soft-ul Quasar.

Capitolul I. Noțiuni teoretice

1.1 ADN

Molecula care conține informații genetice necesare pentru creșterea și dezvoltarea unui organism se numește acid dezoxiribonucleic sau ADN. Este compus din două catene care se rotesc una peste cealaltă și formează structura ADN-ului numită și dublu helix. Coloana vertebrală a fiecărei catene care alcătuiește structura ADN este alcătuită la rândul ei din grupări de fosfat și zahăr (dezoxiriboză) (Figura 1.1). [1]

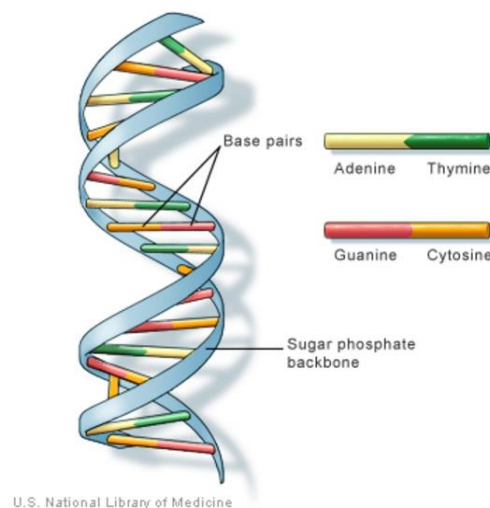


Figura 1.1 Reprezentarea ADN-ului [2]

Adenina asociată cu Timina și Guanina asociată cu Citozina sunt perechile de baze nucleotidice purin-pirimidă care alcătuiesc ADN. Aceste perechi de bază formează legături slabe de hidrogen care țin helixul împreună. [3]

Datorită structurii sale stabile, molecula de ADN poate fi utilizată ca model atât pentru sinteza (transcripție) moleculei ARN (acid ribonucleic), cât și pentru replicarea moleculelor de ADN noi. [3]

Când ADN-ul se divide, se formează două catene unice care funcționează la rândul lor ca șablon pentru sinteza unei catene noi. Noile catene formează și ele legături de hidrogen care formează helixul dublu. Fiecare catenă formată conține o catenă din ADN-ul original și una suplimentară, aceasta replicare folosind la transmiterea genetică. [4]

În celulă, ADN-ul este aranjat în complexe de proteine-ADN numite cromozomi. În celulele eucariote, cromozomii sunt localizați în nucleu, iar ADN-ul se găsește și în mitocondrii și cloroplaste. [4]

În corp se pot produce reacții la nivelul ADN-ului, printre care și metilarea ADN-ului. Metilarea ADN-ului este un proces biologic unde se adaugă grupări de metil în moleculă sau proteine. De exemplu, o genă poate deveni inactivă și poate înceta producerea de proteine dacă o anumită secvență de ADN este metilată. Aceasta reacție de metilare poate produce schimbări a genelor și dezvoltă un risc de apariție a cancerului la persoane. [5]

Cea mai studiată modificarea a ADN-ului este metilarea citozinei (Figura 1.2) care poate afecta accesibilitatea regiunilor genomice la complexe de proteine care influențează structura cromatinei sau afectează rata de transcripție a genei. Un exemplu în acest sens este trecerea celulelor de la o stare sănătoasă la neoplasme maligne în timpul cancerului [6]

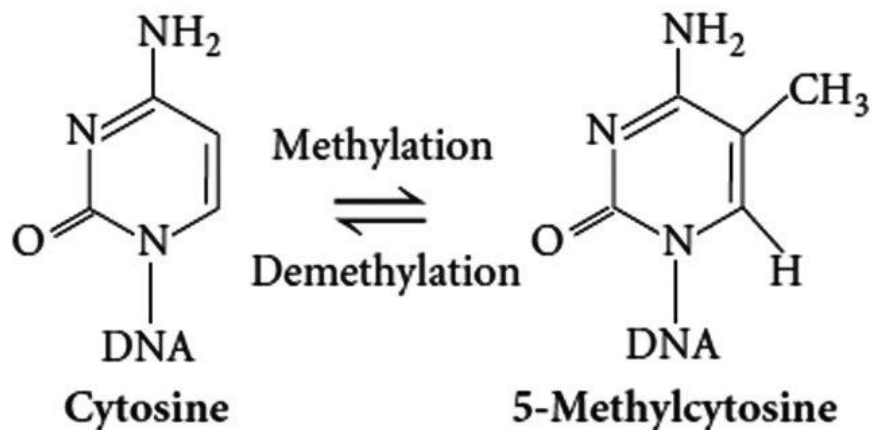


Figura 1.2 Metilarea citozinei [7]

Din studiile recente, s-a demonstrat faptul că metilarea ADN-ului joacă un rol important în inițierea cancerului în organism, susținerea lui și progresia acestuia în organism. [8]

ADN-ul celulelor canceroase au un model de metilare diferit atât cantitativ, cât și calitativ față de ADN-ul celulelor sănătoase. [8] Peisajul metilării modificat epigenic se dovedește a fi o trăsătură comună prezentă în majoritatea tipurilor de cancer și poate servi ca un biomarker universal pentru cancer deoarece apar biomarkeri diferiți pentru fiecare tip de cancer. [9]

Noile cercetări au arătat faptul că metilarea poate avea un efect asupra unor caracteristici specifice ale ADN-ului în soluție, cum ar fi structura sa, flexibilitatea și conformația tridimensională. Acest lucru a fost pus pe seama naturii hidrofobe și mărimilor mai mare ale metilcitozinelor individuale decât citozinele normale. [9]

Unele studii au arătat că ADN-ul genomic obținut din celulele normale prezintă o tendință mai mare de a se agrega în soluții apoase în comparație cu ADN-ul genomic obținut din celulele canceroase. Acest lucru poate din cauza caracteristicilor hidrofobe ale metilcitozinelor care, în funcție de cantitățile lor și de modelul lor, indiferent dacă sunt dispersate uniform sau grupate în grupuri de-a lungul genomului, acestea pot conduce la conformații distincte ale polimerilor ADN în soluție. [9]

Detectia ADN-ului și analizarea sa au jucat un rol important în diferite domenii, cum ar fi prevenția bolilor, diagnostic, tratament și dezvoltarea pastilelor. Până în ultimii ani s-au folosit diferite tehnici de detecție și analiza a ADN-ului cum ar fi: hibridizarea moleculară, reacția în lanț a polimerazei (PCR) și cromatografia lichidă de înaltă performanță-spectroscopie de masă (HPLC-MS). Toate tehnicile tradiționale folosite până acum sunt constrânse de costuri mari, sensibilitate slabă și proceduri complicate, ceea ce duce la pierderea de informații epigenetice ale ADN-ului și face mai greu de identificat întregul ADN. [10]

1.2 SERS (Spectroscopia Raman amplificată de suprafață)

Spectroscopia Raman analizează fotonii împrăștiați din moleculă folosind o singură lungime de undă ca radiație incidentă. Atunci când norul de electroni care înconjoară nucleul este distorsionat de fotonii din radiație incidentă, polarizarea moleculei se schimbă. Atunci când polarizarea unei molecule variază în funcție de vibrația sa, se poate observa efectul Raman. În timpul vibrațiilor legăturilor moleculare, pentru a avea efectul de împrăștiere Raman, variația polarizabilității trebuie să fie diferită de 0. Polarizarea este energia necesară pentru a determina o undă electromagnetică să distorsioneze norul electronic al unei molecule și să producă un moment dipol. [11]

$$\mu = \alpha E \quad (1)$$

α – polarizabilitatea moleculei

Polarizabilitatea moleculei aflată sub un câmp electric cu frecvența ν_0 , poate fi scrisă în serie Taylor.

$$E = E_0 \cos(2\pi\nu_0 t) \quad (2)$$

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q}\right)_0 q + \dots \quad (3)$$

α_0 – polarizabilitatea moleculei în starea de echilibru

R_e, q – deplasarea față de poziția de echilibru

$$q = R - R_e = \alpha \cos(2\pi\nu_i t) \quad (4)$$

ν_i – frecvența vibrației moleculei ca urmare a radiației laser

În concluzie, momentul de dipol indus se poate scrie în funcție de 3 termeni (împrăștiere Stokes și anti-Stokes și radiația Rayleigh)

$$\mu = \alpha_0 E_0 \cos(2\pi\nu_i t) + \frac{1}{2} E_0 \alpha \left(\frac{\partial\alpha}{\partial q}\right)_0 [\cos 2\pi(\nu_i - \nu_0)t + \cos 2\pi(\nu_i + \nu_0)t] \quad (5)$$

Împrăștierea Stokes reprezintă pierderea energiei a fotonilor radiației care ajung pe probă. Aceștia cedează moleculei energia pierdută pentru a fi excitată din nivelul de bază pe primul nivel vibrațional excitat. Împrăștierea anti-Stokes reprezintă fotonii care primesc energie de la moleculele care trec din nivelul vibrațional excitat în nivelul de bază. Radiația Rayleigh reprezintă împrăștierea elastică a fotonilor pe probă (proces întâlnit la cei mai mulți fotoni), proces care nu oferă informații structurale despre moleculele aflate în probă. (Figura 1.3) [12]

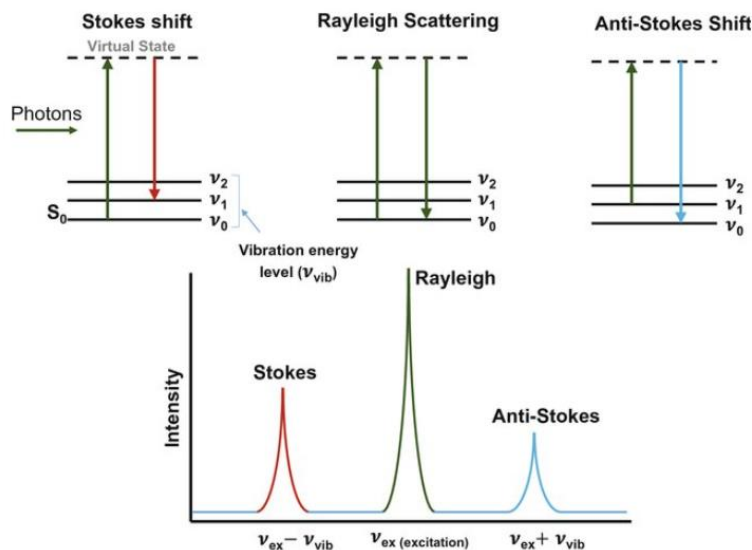


Figura 1.3 Reprezentarea tranzițiilor electronice în spectroscopia Raman [13]

Pe baza împrăștierei inelastice a fotonilor laser, spectroscopia Raman este un tip de spectroscopie vibrațională care poate dezvălui detalii despre structura moleculară a unei probe. Avantajul spectroscopiei Raman este capacitatea de a analiza o probă fără sau cu puțini pași de pre-procesare și primirea rezultatelor rapid. În spectroscopia Raman, doar 1 din 10^6 dintre fotoni sunt împrăștiți inelastic pe probă, ceea ce duce la o sensibilitate scăzută a spectroscopiei. Pentru a putea îmbunătăți sensibilitatea și detectarea de probe mai mici, este nevoie de o metodă ce poate amplifica semnalul spectroscopiei Raman. Una dintre tehnicile care poate îmbunătăți semnalul este SERS. [14]

Pentru efectul de amplificare a spectroscopiei Raman, se pot folosi 2 teorii: teoria electromagnetică și teoria chimică. În ambele teorii este necesară folosirea structurilor metalice de mărimi nanometrice care pot fi absorbite de probă sau aflate în apropierea probei. [15]

Teoria electromagnetică reprezintă o caracteristică a substratului care nu ia în considerare molecula țintă și apare ca urmare a localizării la suprafața substratului a luminii. Factorul de amplificare electromagnetică poate ajunge până la valori de 10^{10} . În teoria electromagnetică, nanostructurile trebuie să fie în apropierea moleculei, în zona de acțiune a câmpului electromagnetic generat de vibrația plasmonică. [15]

În teoria chimică este necesară o interacțiune fizico-chimică între moleculă și substratul metalic. Teoria chimică este dependentă de nivele energetice ale moleculei și de tipul moleculei. Factorul de amplificare chimic poate fi de ordinul 10^3 sau 10^4 . În efectul chimic se întâlnesc două tipuri de interacțiuni: chemioabsorbția și fizioabsorbția (constă într-o interacțiune van de Waals între substratul metalic și moleculă fără a schimba nivelurile energetice ale moleculei.)

Efectul chimic constă în transferul unei sarcini electrice între substratul metalic și molecula absorbită la suprafața metalică. Presupune transferul electronic între nivelurile HOMO și LUMO. HOMO este cel mai înalt orbital molecular ocupat și LUMO este cel mai jos orbital molecular neocupat. [16]

Între împrăștierea Raman și SERS există o relație dată de ecuația:

$$P_{SERS} = G_{SERS} P_{Raman} = G_{SERS}^{EM} G_{SERS}^{Chem} P_{Raman} \quad (6) \quad [17]$$

P_{SERS} – puterea Raman

P_{RAMAN} – puterea SERS

G_{SERS} – factorul de amplificare SERS care rezultă din două mecanisme: amplificarea electromagnetică și amplificarea chimică

$G_{\text{SERS}}^{\text{EM}}$ – factorul de amplificare electromagnetică

$G_{\text{SERS}}^{\text{Chem}}$ – factorul de amplificare chimică [17]

SERS este o metodă analitică care necesită absorbția analiților pe nanosubstraturi metalice (de obicei aur sau argint) care pot fi de diferite forme, cum ar fi nanoparticule sferice, nanostele sau nanodiscuri. Datorită amplificării semnalului Raman, tehnica SERS poate fi utilizată pentru a studia și a detecta cantități mici de substanță, ceea ce o face o tehnică bună pentru numeroase aplicații. [18]

Nanostructurile din metal plasmonic sunt utilizate în tehnicile SERS deoarece ele produc câmpuri electromagnetice intense localizate pe suprafața lor atunci când sunt excitate de un laser ce cade în rezonanță cu oscilația plasmonilor de suprafață. Acest efect duce la o detecție sensibilă, fiind detectate chiar concentrații la nivel de o singură moleculă. [19]

Deoarece moleculele de acid nucleic au semnale Raman intrinseci foarte slabe, care necesită frecvent concentrații extrem de mari pentru a obține semnale fiabile ale moleculelor de ADN, este dificil să se obțină semnalele moleculelor de ADN folosind spectroscopia Raman, care poate oferi o amprentare specifică a diferitelor biomolecule la nivel molecular sau inferior. Datorită acestor probleme întâmpinate în detecția ADN-ului, SERS a deschis noi căi de detectare a structurilor moleculare ale acidului nucleic prin utilizarea nanoparticulelor [18] (Figura 1.4) [20]

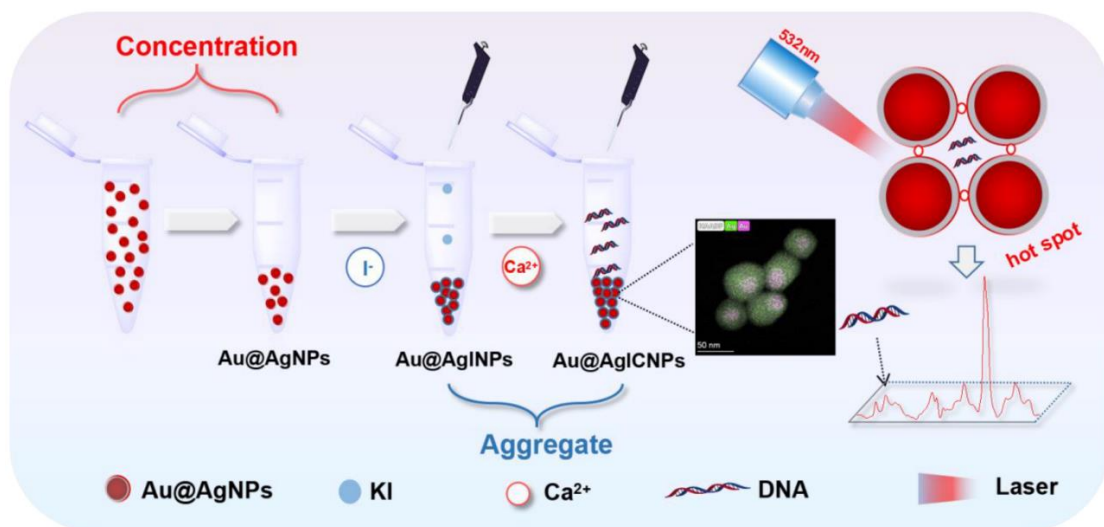


Figura 1.4 Diagrama pentru detecția ADN-ului prin SERS cu nanoparticule [20]

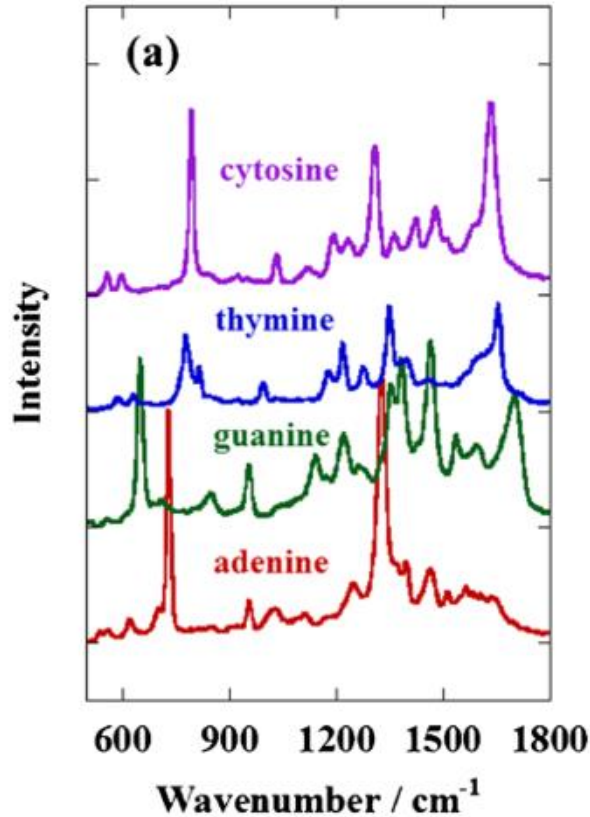


Figura 1.4 Spectrele SERS ale celor 4 nucleotide prezente în ADN (citozină, timină, guanină și adenină) [21]

Spectrul SERS al ADN-ului este dominat de benzile SERS caracteristice bazelor azotate, respective banda la 790 cm^{-1} caracteristică citozei, banda de la 730 cm^{-1} caracteristică adeninei, banda la 640 cm^{-1} caracteristică guaninei și banda de la 770 cm^{-1} caracteristică timinei.

1.3 Leucemie acută promielocitară

Leucemia promielocitară acută este un subtip de leucemie mieloidă acută, adică un cancer cu o creștere rapidă a celulelor hematopoietice din măduva osoasă (reprezintă cam 10% din cazurile de leucemii acute mieloide) și prezintă rate semnificative de vindecare a tratamentului. [22]

În general, leucemia promielocitară acută apare la persoanele de vârstă mijlocie, cu o vârstă medie de 47 de ani. Înainte de 20 de ani, leucemia promielocitară acută este destul de neobișnuită și incidența este mai mare la bărbați decât la femei. [23]

Proteina de fuziune PML-RARa reprezintă produsul unei translocării cromozomiale echilibrate t(15;17) și este prezentă în peste 95% dintre cazurile de leucemii acute promielocitare. Această translocare cromozomială suprapune gena de pe cromozomul 15 PML (gena leucemiei

promielocitare) și gena RAR α (receptor de acid retinoic α) de pe cromozomul 17 (Figura 1.5). [17] Pe brațul lung al cromozomului 17, gena RAR α (acid retinoic α) codifică factorii de transcripție ai receptorilor hormonal nucleari, care sunt implicați întotdeauna în leucemia promielocitară acută (LPA). Odată ce se leagă de acidul retinoic, crește expresia altor gene. [23]

Prezența promielocitelor atipice mari și a precursorilor mieloizi în diferite stadii de dezvoltare în sângele periferic este o caracteristică a leucemiei promielocitare acute. Pentru identificarea în timp util a acestei boli, frunziul de sânge periferic și testarea FISH (hibridizarea fluorescentă in situ) pentru fuziunea PML-RAR α trebuie făcute imediat cum se suspectează LPA. [18]. Testarea FISH constă în atașarea la secvența țintă a unei sonde de ADN monocatenar, marcată fluorescent, pe baza complementarității, de o secvență țintă a unui cromozom. Hibridizarea sondei cu ADN celular este vizualizată la microscopul cu fluorescență echipat cu filtre de excitație și emisie, permițând citirea semnalelor specifice zonei țintă. [24]

Pe lângă aceste tehnici de diagnostic, ar trebui să se efectueze și biopsia măduvei osoase, imunofenotiparea și cariotiparea care poate identifica alte anomalii citogenetice coexistente și subtipuri moleculare neobișnuite de LPA. Cariotiparea este o analiza genetică care poate determina dacă o anumită boală/malformație în legătură cu cromozomii poate fi genetică. O altă metodă de diagnostic este și observarea reacției în lanț a reverstranscriptazei-polimerazei (RT-PCR) pentru PML-RAR α ARN. [23]

Detectia unui transcript PML-RAR α se realizează calitativ, prin izolarea și purificarea ARN-ului mesager, urmată de revers transcriere, pentru obținerea matriței de ADN complementar care va fi folosit în etapa de detecție a secvențelor de ADN țintă cu ajutorul unor primeri specifici. [22]

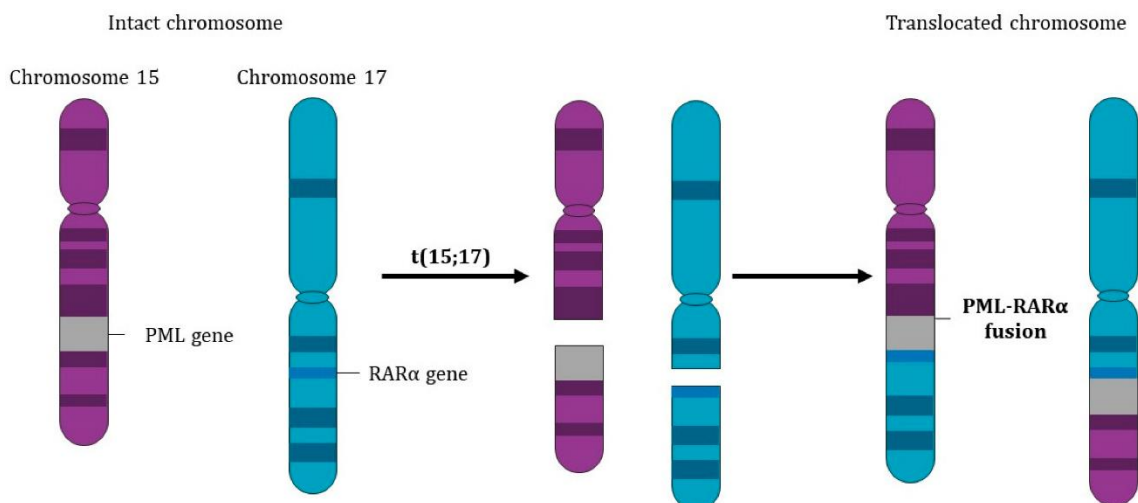


Figura 1.5. Translocatia cromozomiala între cromozomul 15 și cromozomul 17 care rezultă fuziunea genei PML și genei RAR α [25]

Gena PML conține trei puncte de rupere diferite pe cromozomul 15, care sunt situate în interiorul intronului 6 (bcr1 care se întâmplă în 55% din cazuri), exonul 6 (bcr2, care se întâmplă în 5% din cazuri) și intronul 3 (bcr3, care se întâmplă în 40% din cazuri). Deci sunt fezabile trei izoforme PML-RAR α , cunoscute sub numele de long (L sau bcr1), variant (V sau bcr2) și short (S sau bcr3). [26]

Capitolul II. Materiale și metode

2.1. Probele de ADN

Probele de ADN utilizate în această lucrare au fost prelevate de la pacienți diagnosticați cu leucemie promielocitară acută (LPA) și pacienți de control. Pacienții diagnosticați cu LPA sunt în număr de 50, dintre care 31 de pacienți negativi la gena PML-RAR α și 19 pacienți pozitivi la gena PML-RAR α . Prezența genei PML-RAR α a fost detectată prin next generation sequencing (NGS). Pacienții de control sunt în număr de 29.

2.2. Extragerea ADN-ului

Extragerea ADN-ului din sângele pacienților s-a realizat utilizând un mini kit PureLink (Figura 2.1) pentru ADN genomic procesul durând aproximativ 10-15 min. În recipientul de centrifugare se adaugă proba de sânge (640 μ l) preparat cu PureLink Genomic Lysis/Buffer (substanță ce are rolul de a liza conținutul din sânge) și etanol. Recipientul de centrifugare conține un filtru de cu silica, pe care se leagă ADN-ul din lizat. În urma centrifugării ADN-ul rămâne legat de suprafața de silica, restul soluției trece prin filtru. Se centrifughează recipientul la 10.000xg timp de un minut la temperatura camerei. Se elimină tubul de colectare și se introduce recipientul într-un PureLink Collection Tube care se află în kit. Se adaugă 500 μ l de wash buffer (soluție care se folosește pentru a elimina proteinele, ARN-ul și alte componente celulare din ADN) preparat cu etanol în recipient și se centrifugă la 10.000 x g timp de 1 minut la temperatura camerei. După pasul de centrifugare, se elimină din nou tubul de colectare și se introduce recipientul într-un PureLink Collection Tube nou. Se adaugă din nou 500 μ l de wash buffer care este preparat cu etanol pentru a elimina toate impuritățile din ADN și se centrifugă la viteză maximă timp de 3 minute la temperatura camerei, iar apoi se elimină tubul de colectare a impurităților. Recipientul se introduce într-un tub de microcentrifugare steril de 1.5 ml, după se adaugă 200 μ l de PureLink Genomic Elution Buffer în recipient și se centrifugă la temperatura camerei timp de 1 minut la viteză maximă și se obține ADN genomic pur. [27]



Figura 2.1. PureLink mini kit pentru extragerea ADN-ului genomic din sânge [28]

2.3 Sinteza de nanoparticule de argint

Substratul metalic folosit pentru a amplifica semnalul Raman al probelor de ADN a fost o soluție coloidală de nanoparticule de argint (AgNPs) reduse cu hidroxilamină hidrocloridă (hya-AgNPs). Sinteza a fost preparată conform metodei Leopold-Lendl. [29] Mai exact, 17 mg de AgNO_3 au fost dizolvate în 90 mL de apă ultrapură prin mixare. 17 mg de clorhidrat de hidroxilamină au fost dizolvate în 8.8 mL de apă ultrapură, după care s-au adăugat 1.2 mL de NaOH cu o concentrație de 1% în soluția de clorhidrat de hidroxilamină. Amestecul rezultat a fost adăugat în soluția de AgNO_3 sub procesul de mixare, schimbându-se la culoare în maro-galben și obținându-se soluția de nanoparticule de argint. (Figura 2.2). La final, soluția de nanoparticule de argint a fost depozitată la temperatura camerei.



Figura 2.2 Soluția coloidală de nanoparticule de argint reduse cu hidroxilamină hidrocloridă.

2.4 Achiziționarea spectrelor SERS ale probelor de ADN

Pentru achiziționarea spectrelor s-a folosit un spectrometrul Raman InVia Raman (Renishaw) (Figura 2.3) echipat cu laser Nd:YAG cu frecvență dublă cu o lungime de undă de 532 nm. Laserul a fost focusat pe probă printr-un obiectiv de 20X (Leica NA 0.4).



Figura 2.3 Spectrometrul InVia Raman (Renishaw) [30]

Pentru a menține concentrația de ADN uniformă, toate probele de ADN au fost diluate la 20 ng/ μ L. Pentru achiziționarea spectrelor ADN-ului, s-au amestecat 5 μ L de hya-AgNPs, 5 μ L de ADN și 1 μ L de Ca(NO₃)₂ (cu o concentrație finală de 5x10⁻³ M Ca²⁺). Din amestecul rezultat, s-a luat o probă de 5 μ L și a fost plasată sub microscop pe o lamelă acoperită cu o folie de aluminiu (Figura 2.4). Fiecare spectru reprezintă o medie a 3 achiziții de spectre, fiecare de câte 10 secunde.

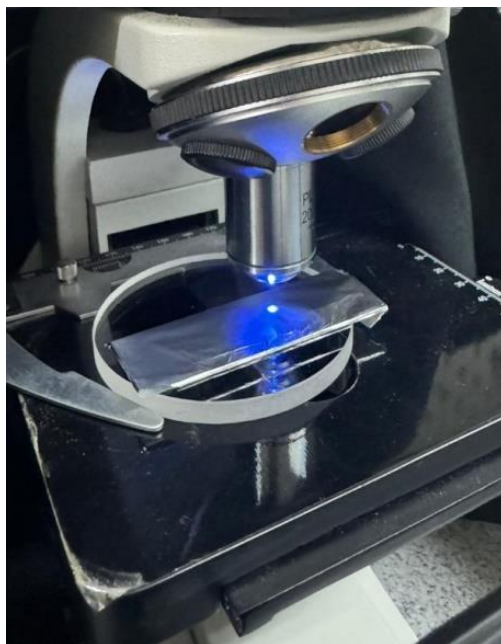


Figura 2.4 Proba de ADN sub microscop

Spectrele de emisie fluorescentă au fost achiziționate folosind sistemul Renishaw, prezentat mai sus. În acest sens, s-a utilizat laserul 445 nm, ce cade în banda absorbție a SybGreen, fluorofor utilizat pentru detecția specifică a ADN-ului dublu catenar. Probele au fost amestecate cu SybGreen 1X, înaintea achiziționării spectrelor de emisie fluorescentă. Pentru achiziționarea spectrelor de fluorescență a timpul de integrare a fost setat la 1 secundă.

2.5 Procesarea spectrelor

Spectrele SERS au fost procesate folosind soft-ul Quasar-Orange [31][32]. A fost păstrată regiunea spectrală 400-1800 cm^{-1} , deoarece benzile SERS ale ADN-ului se află în această regiune, după cum se observă în Figura 1.4. Ulterior, a fost redus baseline-ul spectrelor folosind metoda Rubber-band. Rubber-band este o metodă de reducere a baseline-ului folosind mai multe funcții liniare. Următorul pas a fost normalizarea spectrelor folosind funcția vector normalization.

Scopul final al lucrării este dezvoltarea unor modele de clasificare de tip machine learning (ML) aplicate pe spectrele SERS ale ADN-ului pentru a diferenția probele de AND ce au activat gena PML-RAR α și probele de AND în care această genă nu este activă. Întrucât numărul de probe (50 probe) este semnificativ mai mic decât numărul de variabile (1400 variabile), modelele de clasificare de tip ML sunt susceptibile la overfitting. Pentru a reduce acest risc, înaintea construirii modelelor de clasificare am redus dimensionalitatea setului de date. Reducerea dimensionalității datelor s-a realizat folosind Analiza în componente Principale (PCA). Numărul de Componente Principale (PC) a fost limitat la cele ce au varianța exprimată nenulă. Pentru a ajuta diferențierea celor două grupe (PML-RAR α pozitiv și negativ), a fost testat dacă există diferențe semnificative din punct de vedere statistic între distribuția scorurilor probelor în grupul PML-RAR α pozitiv și negativ pe fiecare Componentă Principală (PC). Acest lucru a fost realizat folosind Student's t-test, cu ipoteza nulă: Nu există diferențe semnificative între media scorurilor pe PC a probelor PML-RAR α pozitiv și negativ. Nivelul de semnificație a fost setat la $p=0.05$.

PC care au rezultat într-o valoare p mai mică de 0.05 au fost utilizate ca variabile input pentru modelele de clasificare ML.

2.6 Modele de clasificare

Machine learning reprezintă algoritmi care învață și face preziceri în baza unor date furnizate, date de „training” fără a primit instrucțiuni explicite. ML este utilizat în diferite domenii, inclusiv în

medicină. În medicină este utilizat pentru diagnostic și tratament. Între metodele de ML se găsesc o serie de modele de clasificare a datelor:

1. K-Nearest Neighbors (kNN) determină clasa unui punct de date pe baza claselor celor mai apropiate k vecine (se folosește pentru seturi de date mici)
2. Random Forest este alcătuit din mai mulți „decision tree” (arbori de decizie) instruiți folosind un set aleatoriu de exemple și atribute de „training”. Adunarea predicțiilor de la mai mulți arbori de decizie minimizează variațiile și se îmbunătățește performanța generală.
3. Naive Bayes este un model bazat pe teorema lui Bayes unde se presupune independența condiționată a caracteristicilor. Fiecare clasă este văzută ca un grup de densități.
4. Logistic Regression este un model liniar în care se folosește funcția logică pentru a prezice probabilitățile claselor binare.

După cum am menționat mai sus, modele de clasificare de tip ML învață un pattern pentru a distinge grupele (pattern dedus pe setul de date training), urmând ca apoi performanța lor să fie demonstrată prin aplicarea aceluși pattern pe un alt set de date (set de date test). Pentru seturile de date cu număr mic de probe, pentru a valida modelul se folosește cross-validare. Astfel, setul de date este împărțit în training și set (procentul de date ce ajung în training și test poate fi ales de utilizator), performanța fiind indicată pe setul de date test. Procedura se repetă de n ori (n putând fi definit de utilizator) iar rezultatul final reprezintă o medie a performanței clasificatorilor pentru toate seturile de date test. Cross-validarea este o tehnică care este folosită pentru evaluarea performanței unui model de clasificare și generalizarea sa. Cross-validarea are ca și scop asigurarea faptului că modelul de clasificare învață într-un mod care nu depinde de un set specific de date și reduce overfitting-ul. Overfitting-ul este un model care se potrivește perfect cu setul de date folosit la training, dar nu și cu un set de date nou.

Înainte de antrenarea modelului se setează hiperparametrii. Hiperparametrii sunt parametrii unui algoritm de ML care influențează modul în care algoritmul poate învăța și pot avea un impact asupra performanței finale. Fiecare model de clasificare are hiperparametrii:

1. În modelul Random Forest există numărul de arbori care se poate modifica
2. În modelul kNN există numărul de vecini folosiți pentru a face o predicție
3. În modelul Logistic Regression există puterea de regularizare care poate fi modificată de la slab la puternic și poate influența modelul.

Capitolul III. Rezultate

3.1. Sinteza AgNPs

Nanoparticulele de Argint utilizate pentru amplificarea semnalului SERS în acest studiu au fost sintetizate conform metodologiei descrise în Capitolul 2, subcapitolul 2.3, metodă introdusă de Leopold și Lendl. [29]. Pentru aceste nanoparticule de argint a fost achiziționat un spectru de absorbție UV-VIS (Figura 3.1), ce arată formarea AgNPs după sinteză. În Figura 3.1 se observă prezența benzii de rezonanță plasmonică a AgNPs, cu un maxim la 408 nm, ce corespunde cu literatura de specialitate [32] pentru AgNPs de formă sferică, acoperită cu ioni de clor, conținuți în soluția de hidroxilamină hidrocloridă.

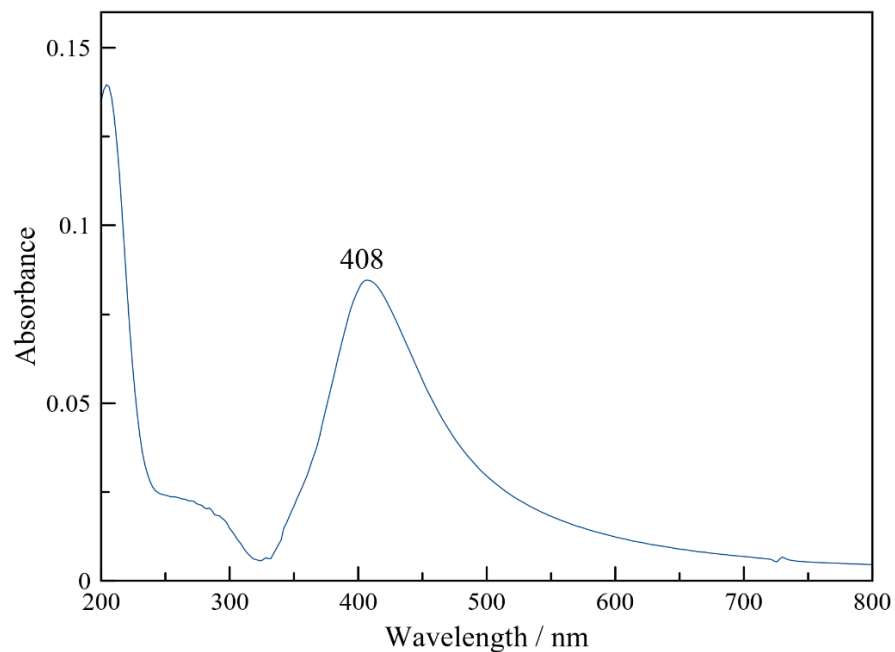


Figura 3.1. Spectrul de absorbție UV-VIS al coloidului hya-AgNPs

Nanoparticulele de Argint au fost caracterizate morfologic și din punct de vedere al stabilității folosind metoda Dynamic Light Scattering, ce a arătat un diametru hidrodinamic mediu al AgNPs de 62 ± 6 nm (Figura 3.2). Stabilitatea hya-AgNPs este demonstrată de Zeta Potențialul de -19 ± 6 mV.

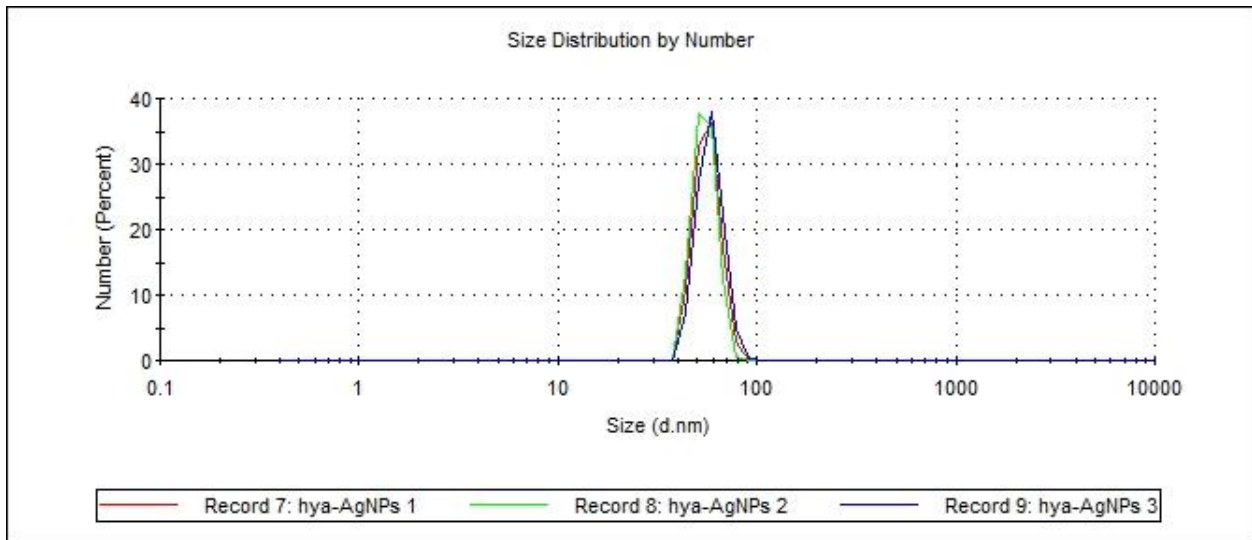


Figura 3.2. Distribuția diametrului hidrodinamic al hya-AgNPs.

3.2. Detectia SERS a ADN-ului

3.2.1. Rolul Ca^{2+} în detectia SERS a ADN-ului

Pentru amplificarea semnalului SERS a ADN-ului, acesta trebuie să se afle în proximitatea nanoparticulelor. În acest sens, Moisoiu et al. [14] utilizează ioni de Ca^{2+} pentru obținerea semnalului SERS a ADN-ului, fără a arăta dacă semnalul ADN-ului nu este deloc vizibil fără adăugarea cationului, sau nu. De aceea, în lucrarea de față am dorit să testez influența Ca^{2+} asupra semnalului SERS a ADN-ului. În toate experimentele Ca^{2+} a fost introdus în soluție sub formă de $\text{Ca}(\text{NO}_3)_2$.

Amestecând ADN cu hya-AgNPs, se observă doar semnalul Raman caracteristic apei (banda de bending H-O-H 1620 cm^{-1}), fără nicio contribuție a ADN-ului. După adăugarea Ca^{2+} se observă apariția semnalului SERS caracteristic ADN (680 cm^{-1} , caracteristic guaninei, 790 cm^{-1} caracteristică citozinei, 730 și 1332 cm^{-1} caracteristic adeninei).

Calciul joacă un rol important în detectia ADN prin SERS deoarece ionii de calciu adăugați ajută la stabilizarea structurii ADN-ului și la atașarea ADN-ului de suprafețele plasmonice. Prin utilizarea calciului se poate îmbunătăți specificitatea tehnicii SERS în detectia ADN-ului prin legarea sa de nanoparticulele de metal.

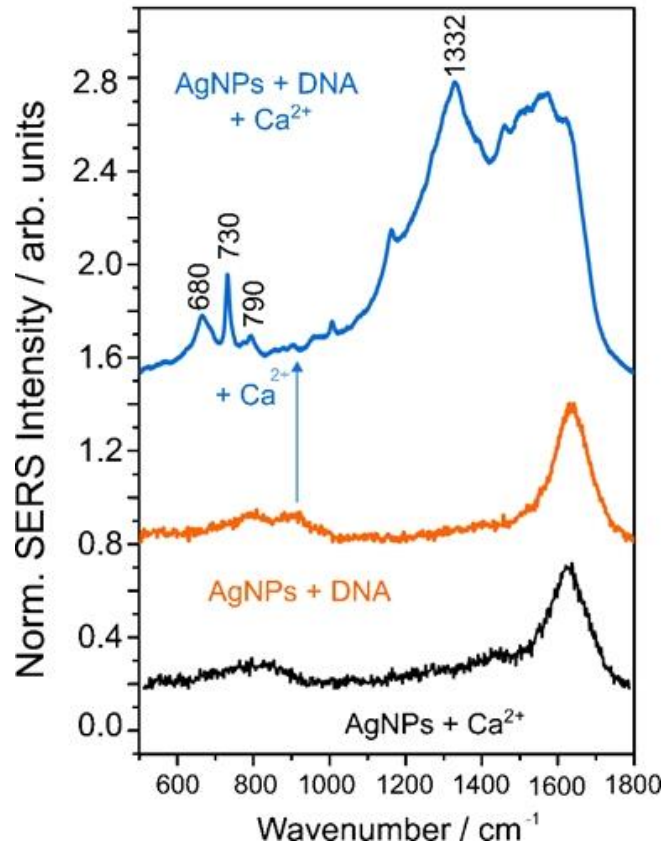


Figura 3.3. Spectrul SERS al ADN-ului în prezența hya-AgNPs înainte și după adăugarea Ca²⁺.

Pentru a vedea dacă amplificarea semnalului SERS a ADN-ului în prezența Ca²⁺ este un rezultat al adsorbției ADN-ului pe suprafața NPs, am achiziționat spectrul de fluorescență al ADN-ului marcat cu SybrGreen înainte și după adăugarea Ca²⁺. (Figura 3.4)

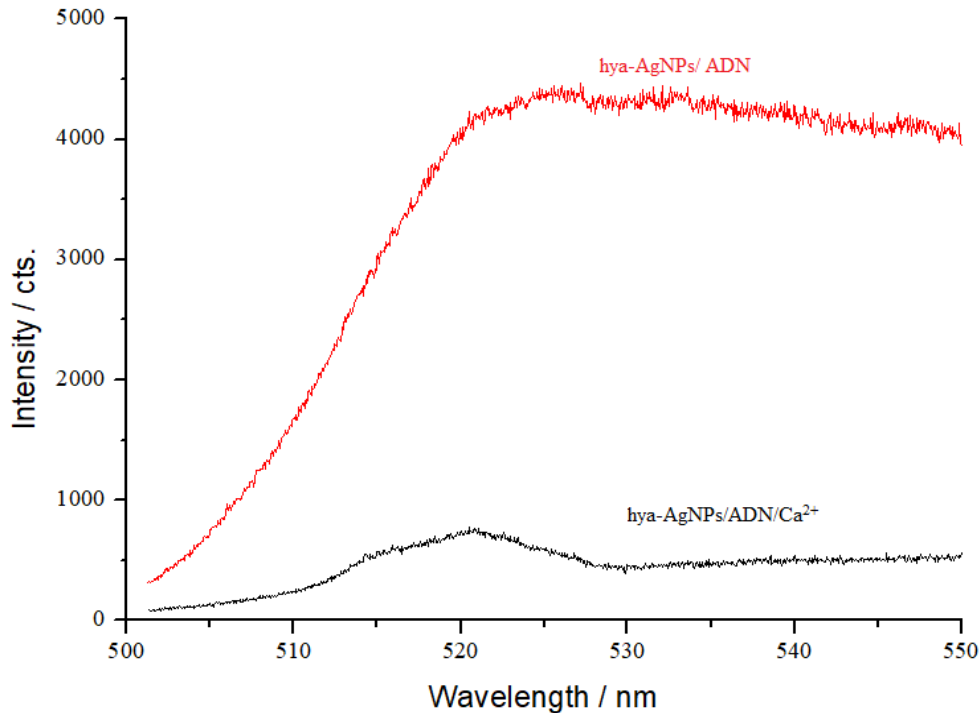


Figura 3.4. Adsorbția ADN pe hya-AgNPs în prezența Ca²⁺. Spectrul de emisie fluorescentă a fluoroforului SybrGreen ce marchează ADN-ul obținut într-o soluție ce conține hya-AgNPs și ADN (roșu) și o soluție ce conține hya-AgNPs/ADN/Ca²⁺ (negru).

Amestecând doar hya-AgNPs cu DNA, ADN nu adsorbe la suprafața AgNPs, ceea ce se observă prin emisia fluorescentă a SybrGreen (maxim la 520 nm, spectru roșu Figura 3.4.) ce se leagă de ADN liber în soluție și emite semnal fluorescent. De menționat este faptul că SybrGreen emite fluorescență doar atunci când se leagă de molecule de ADN. După adăugarea Ca²⁺ (spectru negru Figura 3.4) emisia fluorescență se diminuează ca efect al adsorbției ADN pe suprafața NPs.

Pentru acest studiu numărul total de probe de la pacienți a fost de 114. În timpul achiziției spectrelor s-a observat că pentru anumite probe ADN-ul de era detectabil, iar pentru unele probe era. S-a observat o corelație a raportului absorbției la 260/280, raport ce indică puritatea și calitatea ADN-ului cu apariția sau lipsa spectrului SERS corespunzător ADN-ului. În Figura 3.5 este prezentată media spectrelor SERS pentru probele cu raportul 260/280 mai mică decât 1.8, respectiv cu raportul 260/280 mai mare decât 1.8.

3.2.2. Influența stocării probelor în achiziționarea spectrelor SERS a ADN-ului

ADN-ul analizat inițial a provenit dintr-o bancă de probe adunate pe parcursul a 10 ani. ADN-ul a fost extras ulterior din probele de sânge înghețate și înghețat pentru o perioadă de un an. Ulterior, probele de ADN 20 ng/ μ l au fost măsurate prin spectroscopie SERS. Figura 3.6 prezintă spectrul mediu al probelor de ADN extrase după o înghețare în prealabil a probelor

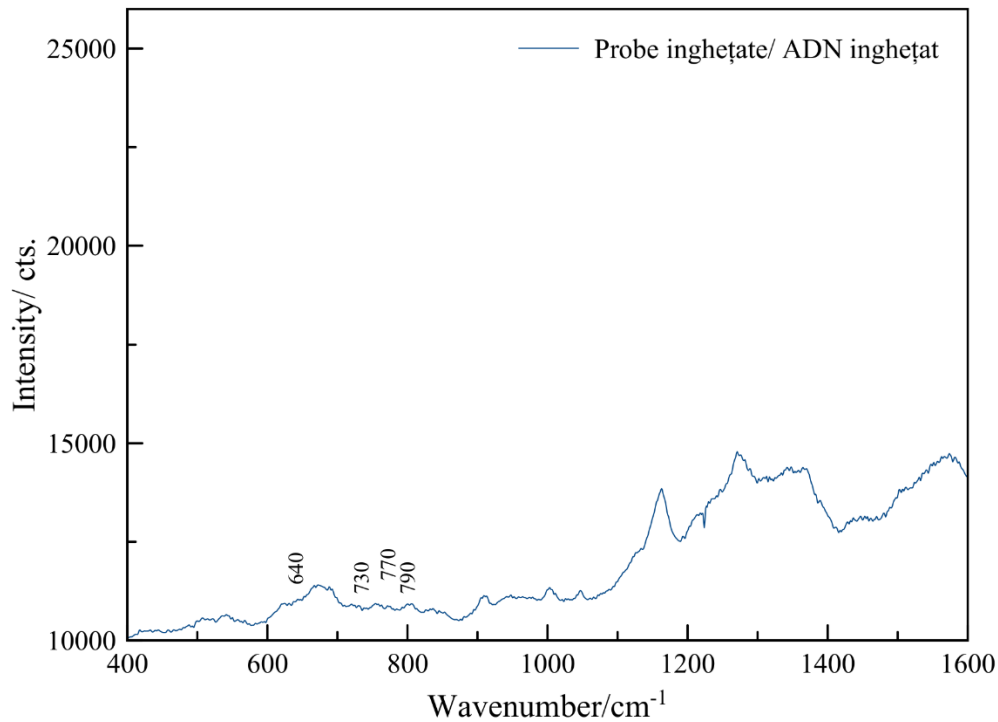


Figura 3.6 Spectrul SERS mediu al probelor de ADN extras din sânge înghețat și stocat. După extracție ADN-ul a fost înghețat și stocat pentru o perioadă de un an înainte achiziționării spectrelor SERS.

După cum se observă în Figura 3.6 în acest caz nu a putut fi înregistrat spectrul SERS al ADN-ului, probabil stocarea acestuia conducând la degradarea lui. Acest lucru se observă și din faptul că mai mult de 50% din probe au avut un raport al Absorbanțelor 260/280 mai mic de 1.8 ceea ce indică impuritatea probelor, probabil ca rezultat a degradării ADN-ului pe perioada stocării.

În continuare, a fost extras ADN-ul din probele de sânge înghețate, urmând ca spectrul SERS al ADN să fie achiziționat fără stocarea prealabilă a acestuia. Figura 3.7 prezintă media spectrelor SERS ale probelor de ADN în acest caz.

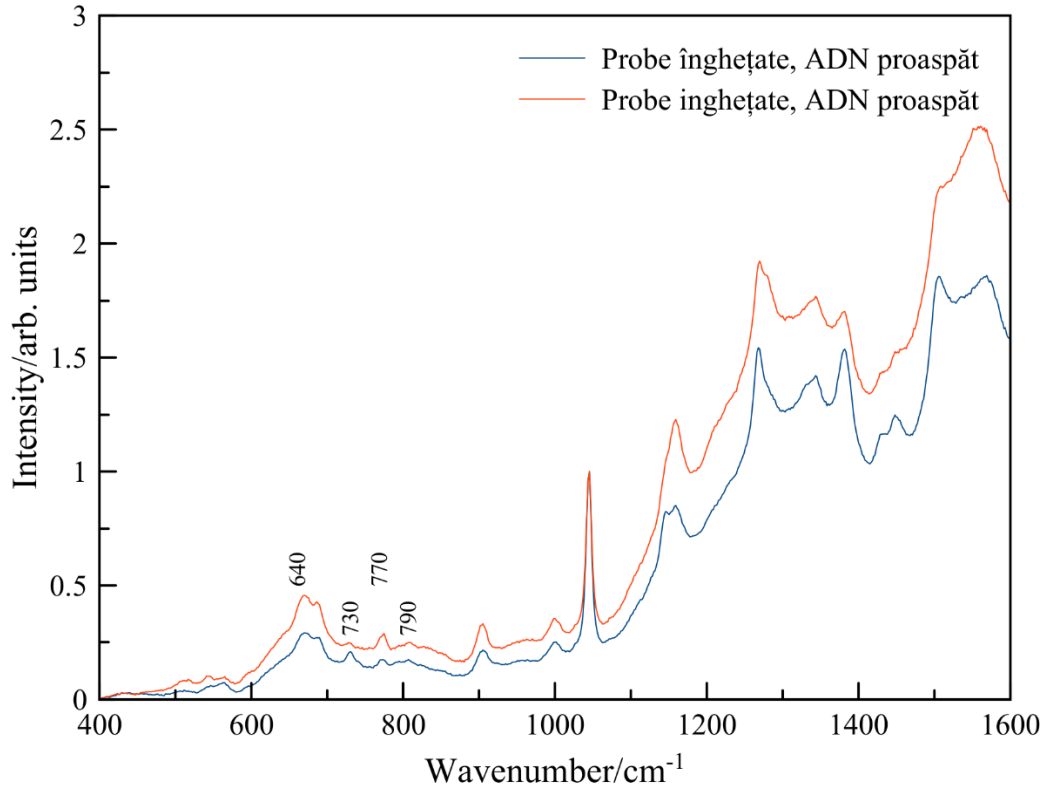


Figura 3.7. Spectrul SERS mediu al probelor de ADN extras din sânge înghețat și stocat. După extracție ADN-ului spectrele SERS au fost înregistrate, fără ca probele să fie stocate.

Benzile SERS caracteristice ADN au fost detectate. Acest lucru arată importanța achiziționării spectrelor SERS imediat după extragerea ADN-ului din probe, stocarea lui pentru o perioadă lungă conducând la imposibilitatea analizei SERS a probelor de ADN.

Pentru a compara dacă stocarea probelor de sânge influențează calitatea ADN-ului și a spectrelor SERS achiziționate, au fost prelevate probe de sânge proaspete de la pacienți control. ADN-ul a fost extras fără înghețarea probelor de sânge. După extragerea ADN-ului au fost achiziționate spectre SERS (Figura 3.8).

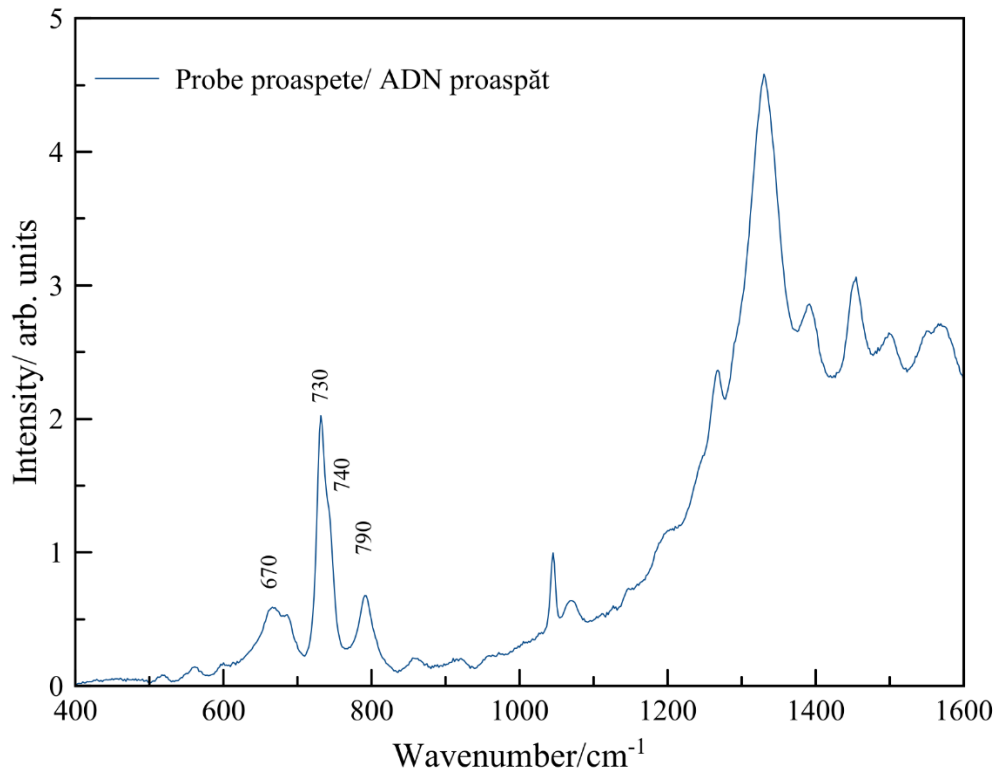


Figura 3.8 Spectrul SERS mediu al probelor de ADN extrase din sânge proaspăt.

Se observă o diferență semnificativă în intensitățile benzilor SERS ale ADN-ului între probele de ADN extrase din probe proaspete de sânge comparative cu cele extrase din probe înghețate și stocate o perioadă lungă de timp. Benzile SERS caracteristice adeninei la 730 și 1330 cm^{-1} , guaninei 680 cm^{-1} și citozinei 790 cm^{-1} sunt semnificativ mai intense în cazul probelor proaspete de sânge. În concluzie, înghețarea și stocarea probelor de ADN extras dar și a probelor de sânge înaintea extracției ADN-ului conduce la diminuarea semnalului SERS. Astfel, pentru a putea utiliza tehnica SERS ca metodă de screening pe baza ADN-ului este important ca probele să fie proaspăt extrase fără înghețare.

ADN-ul din probele control a fost extras din probe proaspete, neînghețate, rezultând într-o calitate mai bună a spectrelor SERS. Acest lucru se observă și prin cantitatea mai mare extrasă, și prin raportul 260/280 de la Nanodrop.

De aceea, o comparație între pacienții CTRL și pacienți diagnosticați cu PML-RAR α nu este posibilă deoarece probele pacienților de control sunt probe proaspete, care nu au fost înghețate și stocate.

3.3. Clasificarea pacienților cu genă PML-RAR α pozitivă și negativă pe baza spectrelor SERS ale ADN-ului

Au fost achiziționate spectre SERS ale ADN-ului extras din probe de sânge de la n=29 CTRL și n=50 pacienți cu cancer, dintre care 31 de pacienți negativi la gena PML-RAR α și 19 pacienți pozitivi la gena PML-RAR α .

După cum am menționat în capitolul 2, subcapitolul 2.5, spectrele au fost procesate în soft-ul Quasar. Workflow-ul realizat pentru procesarea spectrelor și clasificarea lor este prezentat în figura 3.11

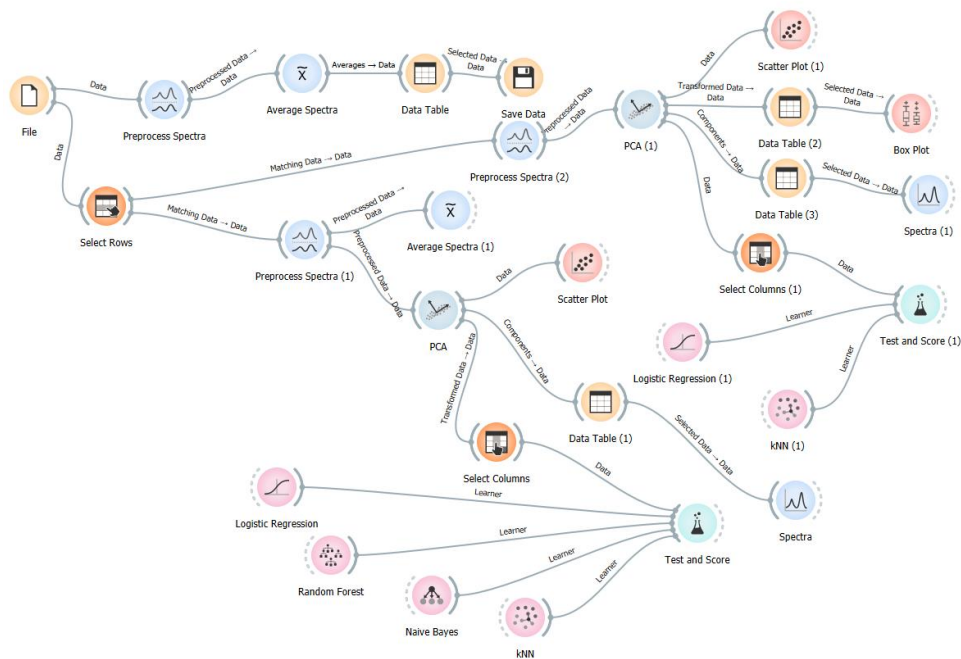


Figura 3.11 Workflow-ul în Quasar utilizat pentru procesarea spectrelor SERS

Spectrele SERS ale probelor de ADN au fost procesate după cum este descris în secțiunea Materiale și metode. Media spectrelor SERS ale ADN-ului cu PML-RAR α pozitiv și negativ este prezentă în figura 3.12.

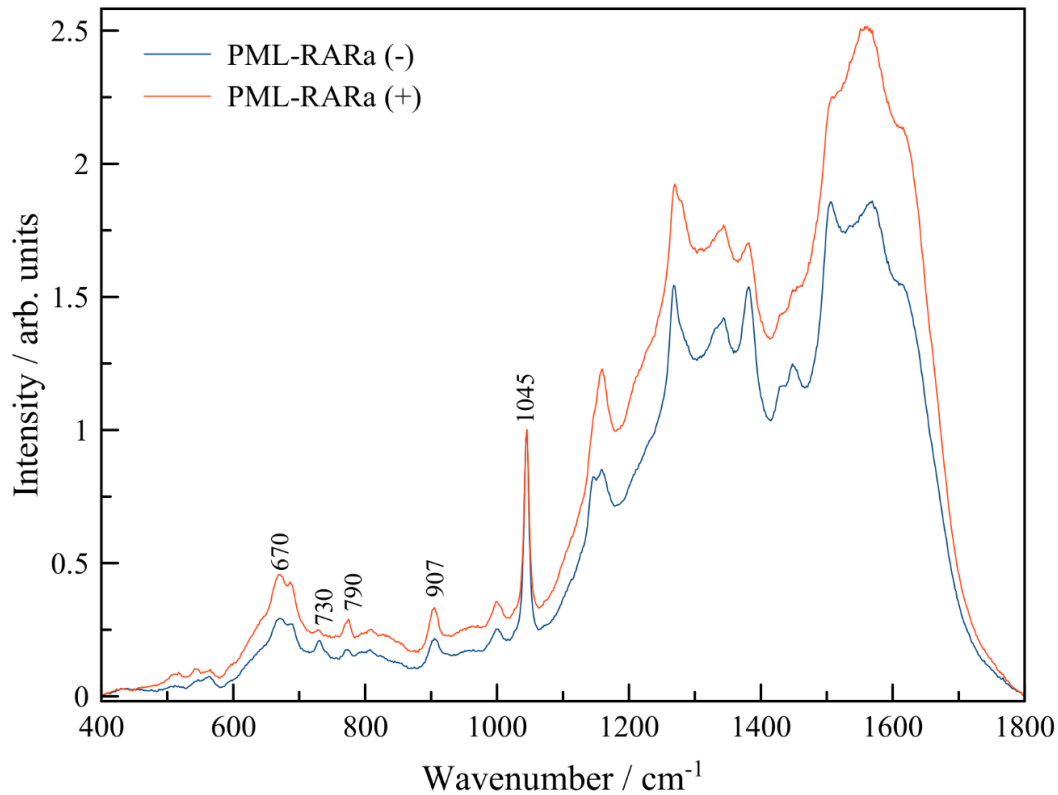


Figura 3.12 Spectrul SERS mediu al probelor de ADN extras de la pacienți cu gena PML-RAR α pozitiv (portocaliu) și negativ (albastru).

Spectrele au fost normalizate la intensitatea benzii Raman 1045 cm⁻¹ caracteristică NO₃, deoarece în toate probele a fost adăugat aceeași cantitate de Ca(NO₃)₂, respectiv concentrație 5x10⁻³ M.

Se observă vizual diferențe între spectrele SERS medii ale probelor de ADN cu gena PML-RAR α pozitivă și negativă. Pentru probele de ADN cu gena PML-RAR α pozitivă banda SERS caracteristică citozinei (790 cm⁻¹) este mai intensă comparativ cu probele unde gena PML-RAR α este negativă. Pe de altă parte, banda SERS caracteristică adeninei (730 cm⁻¹) este mai intensă în cazul lipsei genei PML-RAR α . În plus, banda SERS caracteristică vibrației PO₄⁻ de la backbone-ul ADN-ului este relativ la aceeași intensitate, însemnând că aceeași concentrație de ADN este adsorbită la suprafața hya-AgNPs și detectată prin SERS. Prin urmare, variațiile benzilor SERS caracteristice bazelor azotate, nu sunt datorită unei cantități diferite de ADN adsorbit, ci ca urmare a unei geometrii diferite de adsorbție între probele cu gena PML-RAR α pozitiv și negativ.

În continuare s-a dorit construirea unor clasificatori de tip machine learning care să discrimineze probele de ADN cu gena PML-RAR α pozitivă și negativă pe baza spectrelor SERS ale

ADN-ului. În acest sens, primul pas a fost preprocesarea spectrelor urmată de reducerea dimensionalității setului de date folosind analiza în componente principale (PCA). S-au păstrat primele 3 componente principale (PC) care exprimă 98.8% din varianța inițială. (Figura 3.13).

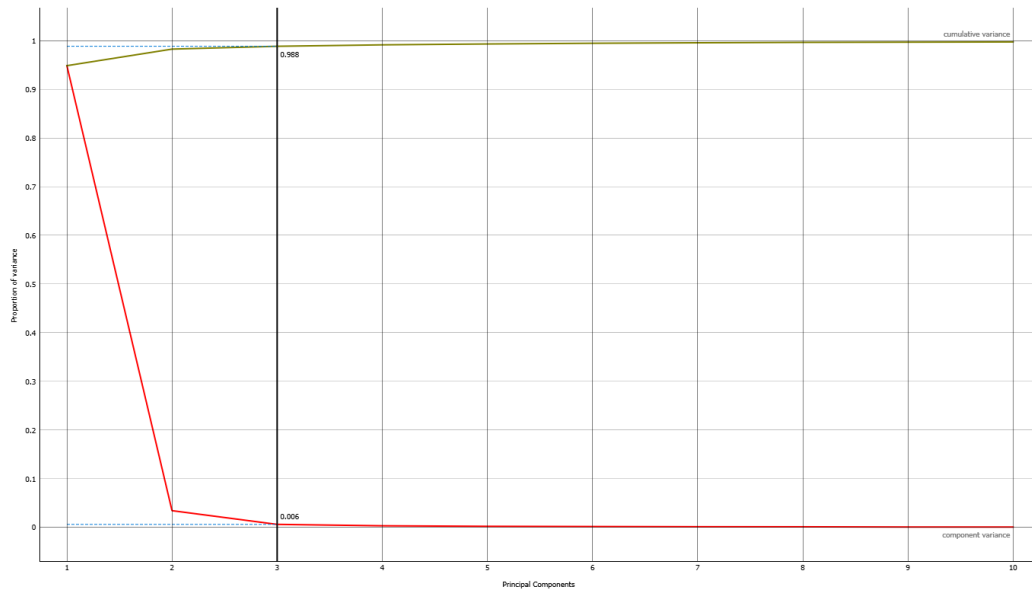


Figura 3.13 Varianța exprimată de primele 10 PC-uri.

Pe baza celor 3 PC-uri a fost urmărită tendința de clusterizare a spectrelor SERS ale ADN-ului în funcție de prezența genei PML-RAR α . Score-ploturile aferente PC1, PC2 respectiv PC3 sunt prezentate în figurile 3.14 și 3.15.

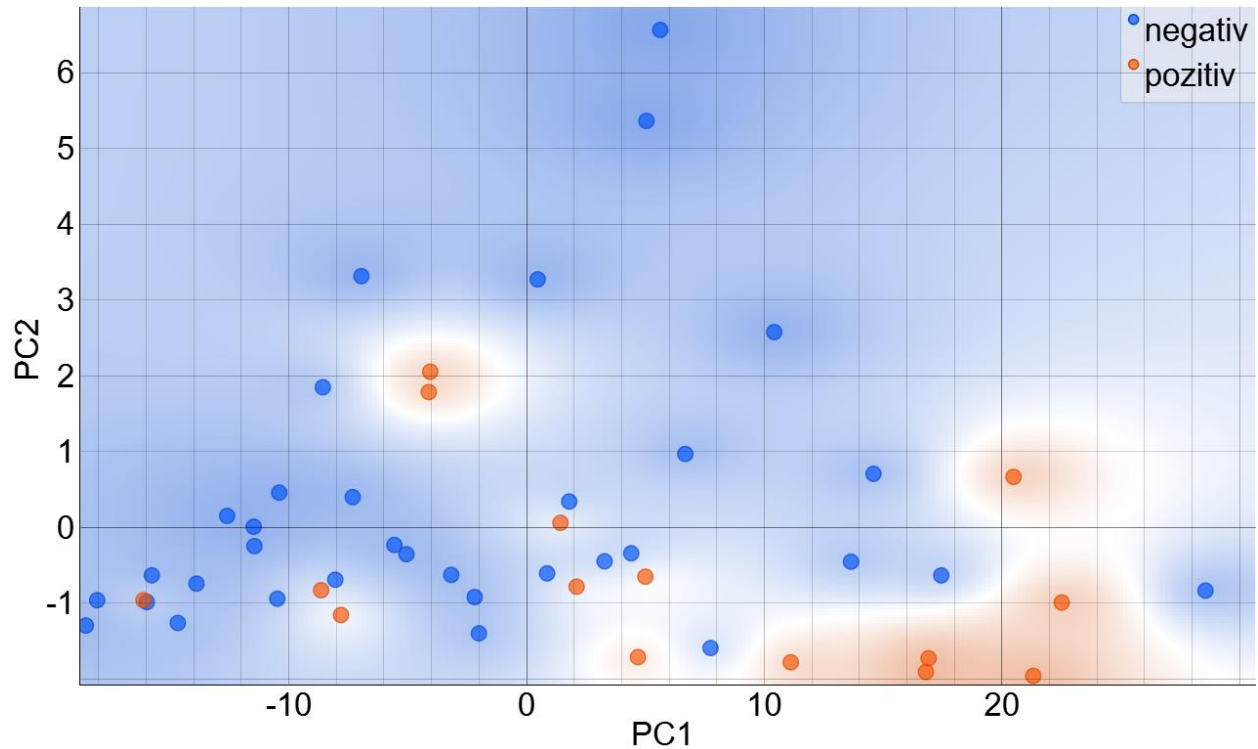


Figura 3.14 Score plot PC1 vs PC2 reprezentând probele de ADN cu gena PML-RAR α negativ (albastru), respectiv pozitiv (portocaliu).

PCA este o metodă nesupervizată de reducere a dimensionalității datelor. Atribuirea culorilor specifice grupului din care face parte fiecare probă se realizează după construirea noilor variabile numite Componente Principale (PC). În acest nou spațiu determinat de PC-uri se poate analiza dacă, în mod nesupervizat, se detectează o informație ce duce la discriminarea datelor.

Se observă o gruparea a probelor cu gena PML-RAR α pozitiv în dreapta jos în scoreplot-ul definit de PC1 vs PC2 și la fel în cazul scoreplot PC1 vs PC3. Probele PML-RAR α negativ sunt împrăștiate pe ambele scoreploturi arătând o variație mai mare a semnalului SERS în interiorul acestei grupe.

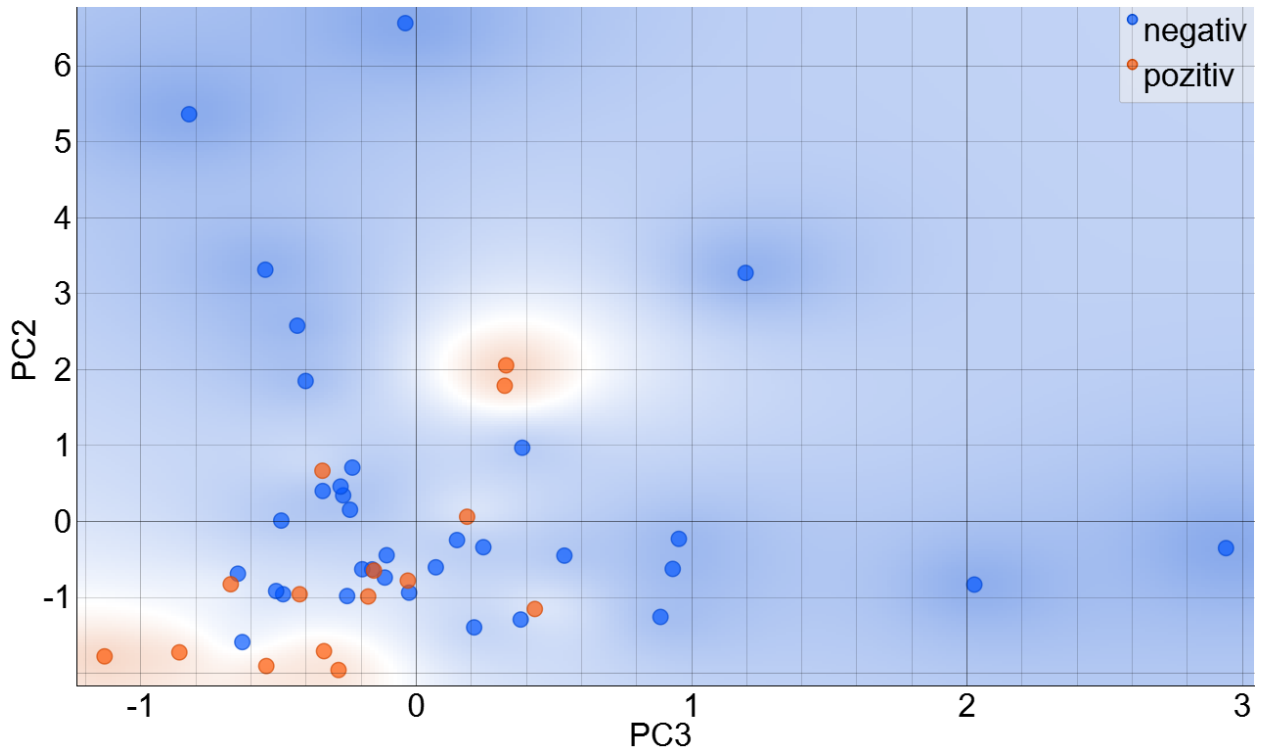


Figura 3.15 Score plot PC2 vs PC3 reprezentând probele de ADN cu gena PML-RAR α negativ (albastru), respectiv pozitiv (portocaliu).

În continuare, am construit diferite modele de clasificare pentru a testa performanța spectroscopiei SERS în detectia genei PML-RAR α . Astfel, scorurile primelor 3 PC-uri au fost folosite ca input pentru modelele de clasificare kNN, logistic regression, random forest și naive bayes.

Toate modelele au fost cross-validate folosind metoda leave-one-out. Hiperparametrii au fost determinați prin trial-and-error pentru a obține acuratețea de clasificare maximă. Performanța modelelor este măsurată prin matrice de performanță cum ar fi: acuratețea, recall, F1-score și specificitatea. Tabelul 3.1 prezintă matricele de performanță pentru clasificarea probelor cu gena PML-RAR α pozitiv, respectiv negativ.

Tabel 3.1. Performanța de clasificare a probelor de ADN cu gena PML-RAR α pozitiv și negativ pe baza spectrelor SERS folosind modelele kNN, Logistic Regression, Naive Bayes și Random Forest

Model	Acuratețea de clasificare	de Scor F1	Sensibilitate	Specificitate
kNN	0.735	0.707	0.735	0.510
LR	0.776	0.752	0.776	0.566
Naive Bayes	0.633	0.633	0.633	0.503
Random Forest	0.735	0.726	0.735	0.585

În urma rezultatelor obținute la modele de clasificare, modelul kNN a rezultat o acuratețe de clasificare de 73%, la modelul Logistic Regression s-a obținut o acuratețe de clasificare de 77%, la modelul Naive Bayes o acuratețe de clasificare de 63% și la modelul Random Forest o acuratețe de clasificare de 73%. Media acurateței probelor de ADN cu gena PML-RAR α pozitive și negative este de 70%.

În continuare, știind că majoritatea benzilor SERS caracteristice bazelor azotate din ADN-ului se află în regiunea 600-1000 cm^{-1} , am folosit pentru clasificare doar această regiune în spectru pentru a vedea dacă putem îmbunătăți modelul. Regiunea din spectru SERS de la 1000-1800 cm^{-1} prezintă benzi caracteristice AND, însă se observă o suprapunere mare a acestora. (Figura 3.16)

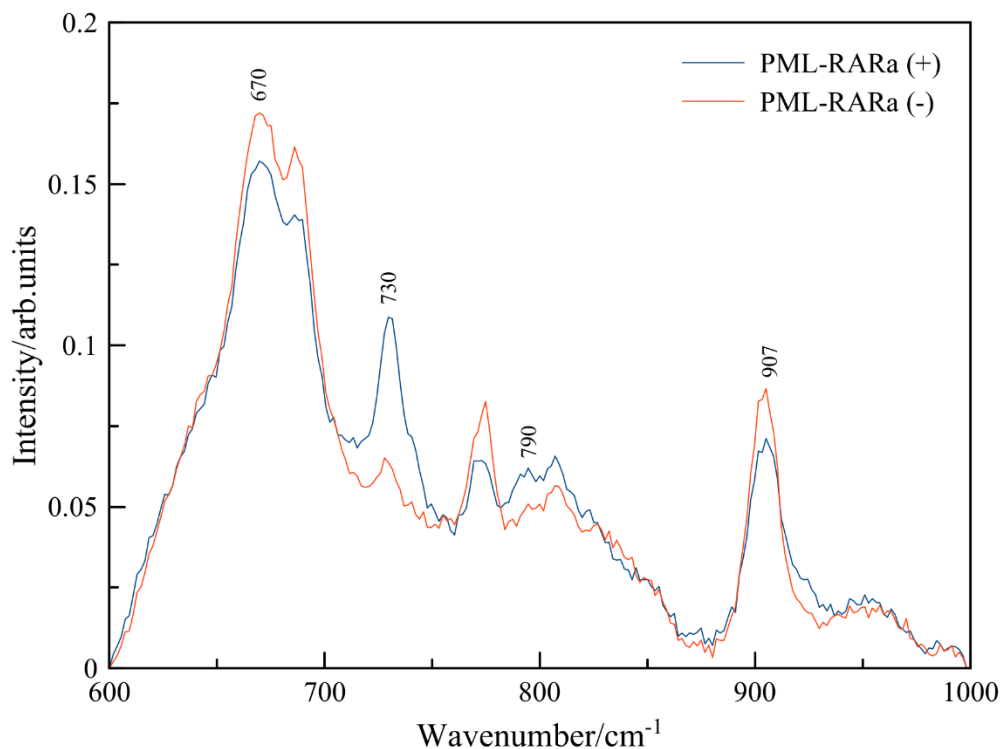


Figura 3.16 Spectrul SERS mediu al probelor de ADN cu gena PML-RAR α pozitivă, respectiv negativă în regiunea spectrală 600-1000 cm⁻¹

Ca o comparație, am procesat spectrele păstrând regiunea spectrală 600-1000 cm⁻¹ și au fost realizați aceiași pași de procesare și modele de clasificare ca mai sus. În acest caz, la analiza PCA au fost păstrate primele 14 PC-uri care exprima 90,3 % din varianță (Figura 3.17).

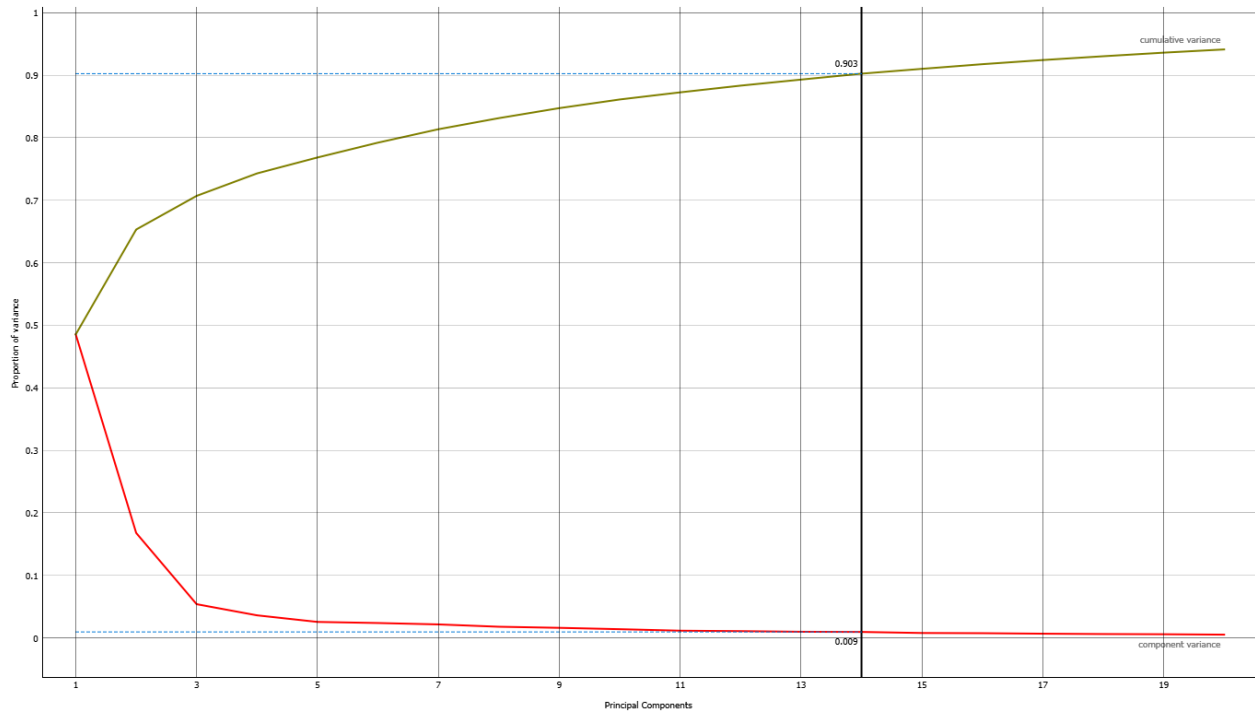


Figura 3.17. Variața exprimată de primele 19 PC-uri.

Cum am menționat anterior PCA este o metodă nesupervizată, ceea ce înseamnă că nu ține cont de grupa din care fiecare probă face parte. Astfel, chiar dacă PC-urile conțin informație spectrală, și nu zgomot, nu este neapărat necesar ca această informație să fie relevantă pentru discriminarea grupelor. Astfel, pentru a ajuta modelele de clasificare, se poate face o filtrare la variabilelor înaintea construirii modelelor de tip ML. În cazul nostru, filtrarea variabilelor am făcut-o utilizând Student's t-test ce testează dacă distribuțiile scorurilor unei variabile se suprapun în cele două grupe, PML-RAR α pozitiv și negativ. Dacă distribuțiile se suprapun, valoare p rezultate în urma testului este mai mare de 0.05 și înseamnă că din punct de vedere statistic acea variabilă nu ajută la discriminarea celor două grupe. Dacă valoare p este mai mică de 0.05 înseamnă că există o diferență între scorurile probelor cu PML-RAR α pozitiv și negativ și putem folosi acea variabilă pentru a clasifica probele, așadar am folosit PC1, PC2 și PC4 (Figura 3.18, Figura 3.19, Figura 3.20)

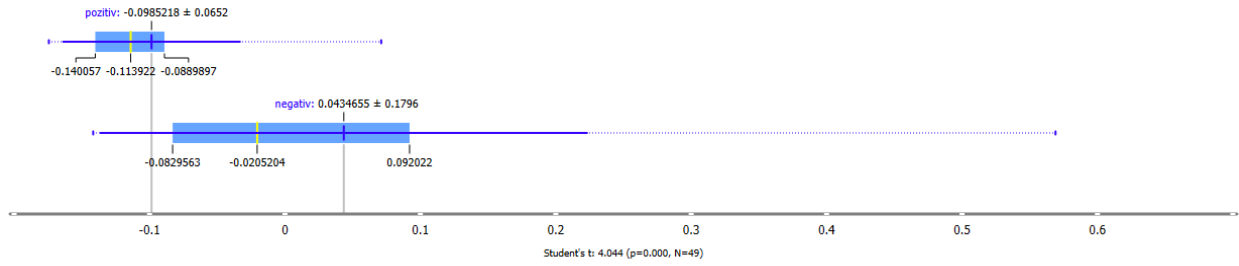


Figura 3.18 Student's t-test pentru PC1

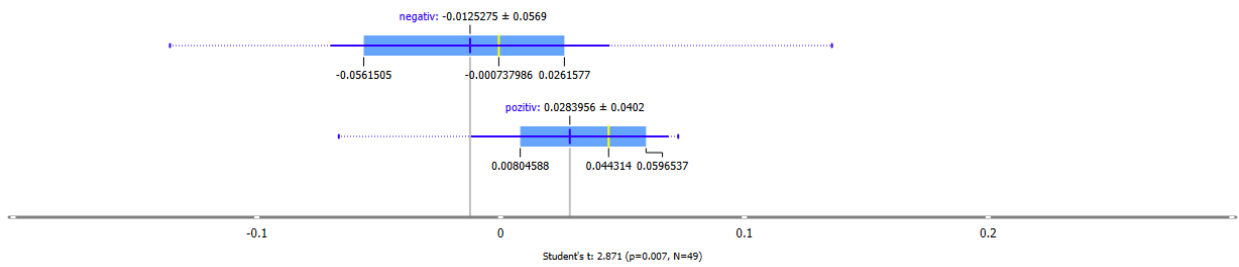


Figura 3.19 Student's t-test pentru PC3

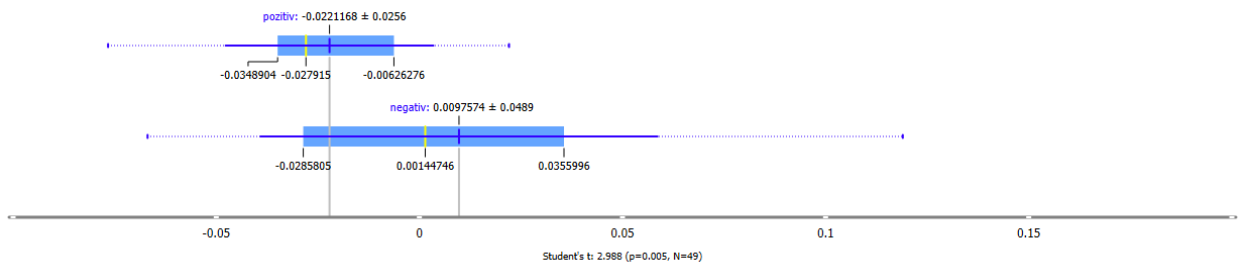


Figura 3.20 Student's t-test pentru PC4

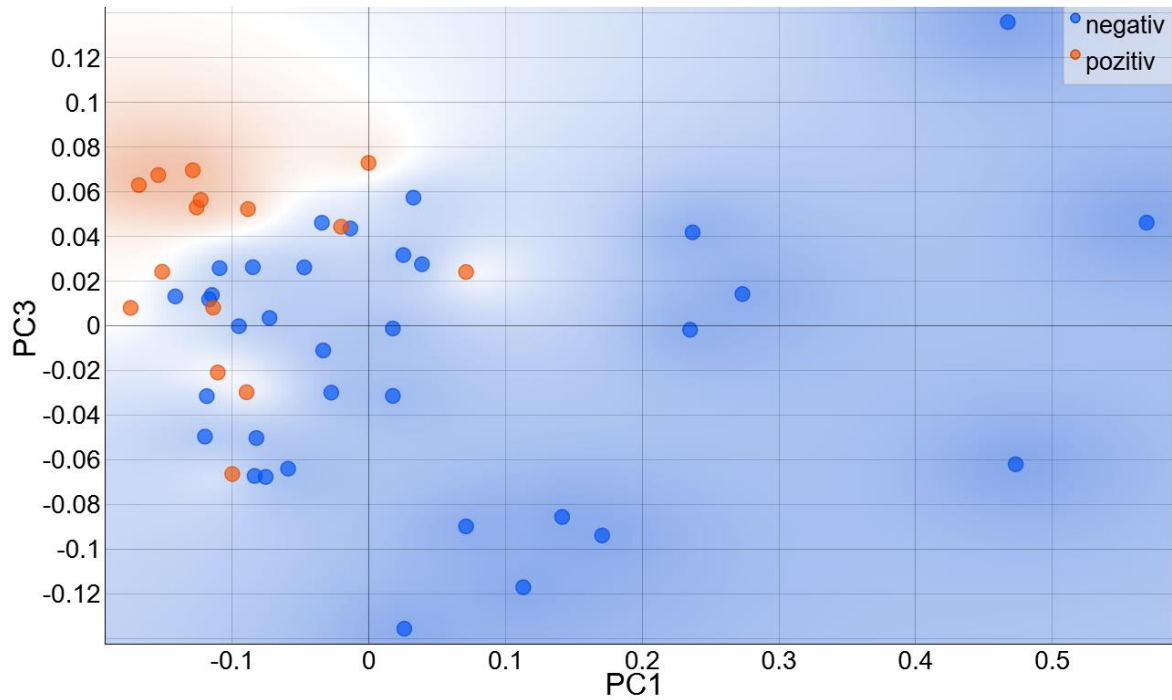


Figura 3.21 Score plot-ul pentru PC1 vs PC2 a probelor de ADN cu gena PML-RAR α pozitivă (portocaliu), respectiv negativă (albastru)

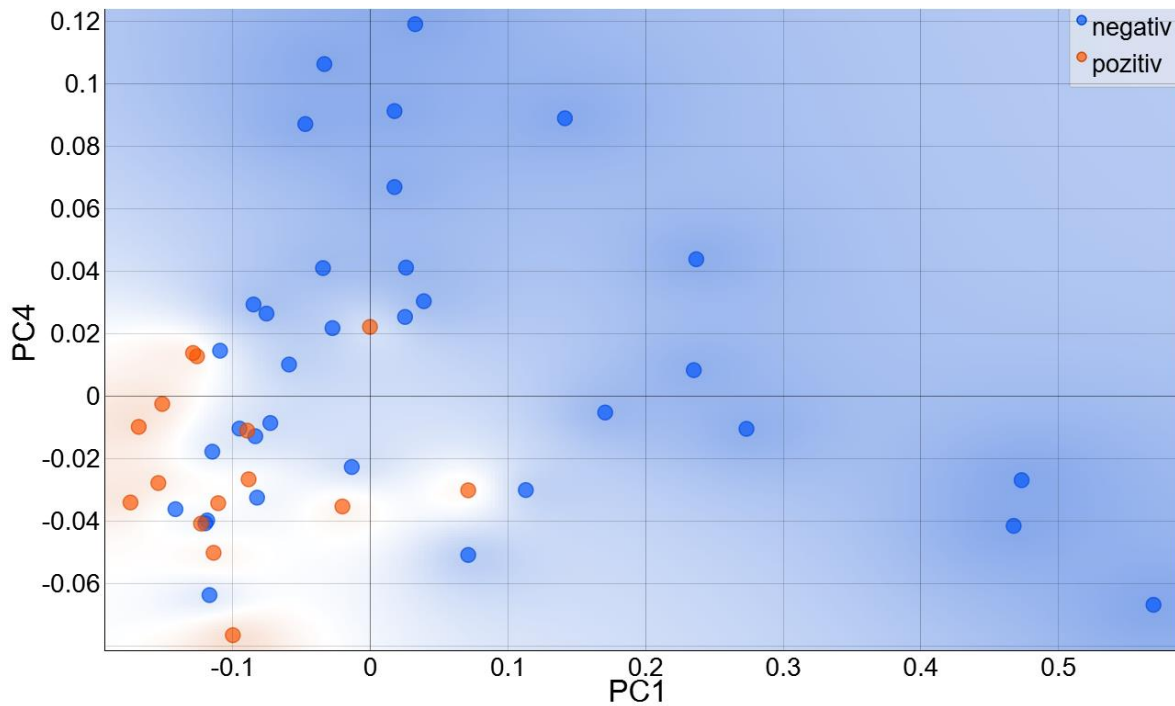


Figura 3.22 Score plot-ul pentru PC1 vs PC4 a probelor de ADN cu gena PML-RAR α pozitivă (portocaliu), respectiv negativă (albastru)

După realizarea testului, am aplicat același pas următor, Test and Score pentru a verifica performanța. Am folosit aceleași modele de clasificare: Logistic Regression, Naive Bayes, kNN și Random Forest, iar rezultatele obținute sunt prezentate în tabelul 3.2.

Tabel 3.2. Performanța de clasificare a probelor de ADN cu gena PML-RAR α pozitiv și negativ pe baza spectrelor SERS folosind modelele kNN, Logistic Regression, Naive Bayes și Random Forest .

Model	Acuratețe	F1	Sensibilitate	Sensitivitate
kNN	0.694	0.568	0.694	0.306
Random Forest	0.735	0.736	0.735	0.585
Naive Bayes	0.653	0.661	0.653	0.586
LR	0.796	0.796	0.796	0.724

În urma rezultatelor obținute la modele de clasificare, modelul kNN a rezultat o acuratețe de clasificare de 69%, la modelul Random Forest s-a obținut o acuratețe de clasificare de 73%, la modelul Naive Bayes o acuratețe de clasificare de 65% și la modelul Logistic Regression o acuratețe de clasificare de 79%. Media acurateței probelor de ADN cu gena PML-RAR α pozitive și negative este de 75%.

Ca și o concluzie pentru rezultatele obținute în urma modelelor de clasificare pentru regiunea spectrală 400-1800 cm^{-1} și regiunea 600-1000 cm^{-1} s-a observat faptul că rezultatele și acuratețea probelor ADN cu gena PML-RAR α depind și de regiunea spectrală aleasă.

Concluzii

În această lucrare s-au achiziționat spectre SERS de la probe de ADN extrase de sânge de la pacienți cu leucemie acuta promioloicitara (n=50), dintre care 31 cu gena PML-RAR α pozitivă, respectiv 19 cu gena PML-RAR α negativă. Pentru achiziționarea spectrelor SERS s-a folosit Ca²⁺ deoarece conduce la adsorbția ADN-ului la suprafața nanoparticulelor de Ag și pentru a amplifica semnalul SERS. S-a observat faptul că stocarea și înghețarea probelor de sânge și ADN-ul extras și stocat poate duce la un semnal slab SERS și degradarea ADN-ului, din acest motiv ADN-ul a fost proaspăt extras din probele de sânge care au fost înghețate și stocate pentru a obține un semnal mai bun al spectrelor SERS. După extragerea ADN-ului din nou s-a observat o îmbunătățire și o diferență a semnalului SERS, așadar cu cât probele de ADN sunt mai proaspete, fără a fi înghețate și stocate, obținem un semnal SERS favorabil.

Metoda actuală de verificare a prezenței genei PML-RAR α se face prin secvențierea ADN-ului, care poate fi o metodă costisitoare, o metodă care necesită mult timp pentru obținerea rezultatelor și o metodă ce necesită o pregătire mai amplă a probelor de ADN extrase. Ca și o concluzie la rezultatele obținute pe baza spectrelor SERS obținute de la probele de ADN cu gena PML-RAR α pozitivă, respectiv negativă se poate determina cu o acuratețe de 79.6%. Studiul de față reprezintă un studiu proof-of-principle ce trebuie validat pe o cohortă mai mare de pacienți cu gena PML-RAR α cunoscută.

Bibliografie

1. <https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid>
2. <https://medlineplus.gov/genetics/understanding/basics/dna/>
3. <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/dna>
4. <https://www.britannica.com/science/DNA>
5. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/methylation>
6. Kumar, S., Chinnusamy, V., & Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00640>
7. Saini, A., Mastana, S., Myers, F., & Lewis, M. (2013). 'From Death, Lead Me to Immortality' – Mantra of Ageing Skeletal Muscle. *Current Genomics*, 14(4), 256–267. <https://doi.org/10.2174/1389202911314040004>
8. Lakshminarasimhan, R., & Liang, G. (2016). *The Role of DNA Methylation in Cancer* (pp. 151–172). https://doi.org/10.1007/978-3-319-43624-1_7
9. Sina, A.A.I., Carrascosa, L.G., Liang, Z. *et al.* Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker. *Nat Commun* 9, 4915 (2018). <https://doi.org/10.1038/s41467-018-07214-w>
10. Zhang, T., Quan, X., Cao, N., Zhang, Z., & Li, Y. (2022). Label-Free Detection of DNA via Surface-Enhanced Raman Spectroscopy Using Au@Ag Nanoparticles. *Nanomaterials*, 12(18), 3119. <https://doi.org/10.3390/nano12183119>
11. Leopold, N. Surface-enhanced Raman spectroscopy, selected applications. Editura Napoca Star
12. J. R. Ferraro, Introductory raman spectroscopy. Elsevier, 2003.
13. Cho, Y. C., & Ahn, S. il. (2020). Fabricating a Raman spectrometer using an optical pickup unit and pulsed power. *Scientific Reports*, 10(1), 11692. <https://doi.org/10.1038/s41598-020-68650-7>
14. Moisoiu, T., Iancu, S. D., Burghilea, D., Dragomir, M. P., Iacob, G., Stefancu, A., Cozan, R. G., Antal, O., Bálint, Z., Muntean, V., Badea, R. I., Licarete, E., Leopold, N., & Elec, F. I. (2022). SERS Liquid Biopsy Profiling of Serum for the Diagnosis of Kidney Cancer. *Biomedicines*, 10(2), 233. <https://doi.org/10.3390/biomedicines10020233>

15. M. Muniz-Miranda, F. Muniz-Miranda, and A. Pedone, "Spectroscopic and computational studies on ligand-capped metal nanoparticles and clusters," *Metal Nanoparticles and Clusters: Advances in Synthesis, Properties and Applications*, pp. 55–87, 2018
16. C. Boerigter, U. Aslam, and S. Linic, "Mechanism of charge transfer from plasmonic nanostructures to chemically attached materials," *ACS nano*, vol. 10, no. 6, pp. 6108–6115, 2016
17. R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, and L. Fabris, "A review on surface-enhanced Raman scattering," *Biosensors*, vol. 9, no. 2, p. 57, 2019
18. John, N., & George, S. (2017). Raman Spectroscopy. In *Spectroscopic Methods for Nanomaterials Characterization* (pp. 95–127). Elsevier. <https://doi.org/10.1016/B978-0-323-46140-5.00005-4>
19. Huang, J.A., Mousavi, M.Z., Zhao, Y. *et al.* SERS discrimination of single DNA bases in single oligonucleotides by electro-plasmonic trapping. *Nat Commun* **10**, 5321 (2019). <https://doi.org/10.1038/s41467-019-13242-x>
20. <https://medlineplus.gov/genetics/understanding/basics/dna/>
21. Fukunaga, Y., Harada, M. & Okada, T. Surface-enhanced Raman scattering of DNA bases using frozen silver nanoparticle dispersion as a platform. *Microchim Acta* **188**, 406 (2021). <https://doi.org/10.1007/s00604-021-05055-x>
22. <https://www.ritusbiotec.ro/detectia-calitativa-a-proteinei-de-fuziune-pml-rar%CE%B1-bcr1-bcr2-bcr3/>
23. Shashank R. Cingam, Nebu V. Koshy, "Acute Promyelocytic Leukemia"
24. <https://geneticlab.ro/articole/testul-fish-pentru-cromozomii-13-18-21-x-y/>
25. de Almeida, T. D., Evangelista, F. C. G., & Sabino, A. de P. (2023). Acute Promyelocytic Leukemia (APL): A Review of the Classic and Emerging Target Therapies towards Molecular Heterogeneity. *Future Pharmacology*, 3(1), 162–179. <https://doi.org/10.3390/futurepharmacol3010012>
26. Behdad, A., & Betz, B. L. (2017). Molecular Testing in Acute Myeloid Leukemia. In *Diagnostic Molecular Pathology* (pp. 419–434). Elsevier. <https://doi.org/10.1016/B978-0-12-800886-7.00033-9>
27. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FSLG%2Fmanuals%2Fpurelink_genomic_mini_man.pdf
28. <https://www.fishersci.co.uk/shop/products/invitrogen-purelink-viral-rna-dna-mini-kit/10013922>

29. Leopold, N., & Lendl, B. (2003). A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. *The Journal of Physical Chemistry B*, 107(24), 5723–5727. <https://doi.org/10.1021/jp027460u>
30. <https://www.directindustry.com/prod/renishaw/product-5200-56003.html>
31. M. Toplak, et al. Quasar: Easy Machine Learning for Biospectroscopy. *Cells* 10, 2300 (2021)
32. Toplak, et al. Infrared Orange: Connecting Hyperspectral Data with Machine Learning. *Synchrotron Rad. News* 30, 40–45 (2017)
33. Ho, N. T., Tien, H. N., Jang, S.-J., Senthilkumar, V., Park, Y. C., Cho, S., & Kim, Y. S. (2016). Enhancement of recombination process using silver and graphene quantum dot embedded intermediate layer for efficient organic tandem cells. *Scientific Reports*, 6(1), 30327. <https://doi.org/10.1038/srep30327>

DECLARAȚIE PE PROPRIE RĂSPUNDERE

Subsemnatul, Havrici Ana Maria declar că Lucrarea de licență/diplomă/disertație pe care o voi prezenta în cadrul examenului de finalizare a studiilor la Facultatea de Fizică, din cadrul Universității Babeș-Bolyai, în sesiunea iulie 2024 , sub îndrumarea Prof. dr. Nicolae Leopold , reprezintă o operă personală. Menționez că nu am plagiat o altă lucrare publicată, prezentată public sau un fișier postat pe Internet. Pentru realizarea lucrării am folosit exclusiv bibliografia prezentată și nu am ascuns nici o altă sursă bibliografică sau fișier electronic pe care să le fi folosit la redactarea lucrării.

Prezenta declarație este parte a lucrării și se anexează la aceasta.

Data,

24.06.2024

Nume,

Havrici Ana Maria

Semnătură