

Raport activitate privind proiectul (PN-II-PT-PCCA-2013-4-0202) nr. 7/2014

SISTEM INTELIGENT DE ASISTARE AUTOMATA ÎN DIAGNOSTICUL CANCERULUI CERVICAL (PAPDIA)

Raport de cercetare GeneticLab

Cancerul colului uterin este o problemă de sănătate publică în întreaga lume, afectând cu precădere femeii între 35-50 ani aflate în perioada activă socială și profesională. Dintre țările Uniunii Europene, România prezintă una dintre cele mai ridicate rate a mortalității prin cancer de col uterin, principalul tip de cancer genital și a doua cauză de deces prin cancer la femeile din țara noastră. Depistarea leziunilor colului uterin într-o etapă timpurie a carcinogenezei și abordarea integrată a informațiilor clinice și citologice, cu modularea conduitei diagnostice și terapeutice conform unor protocoale larg acceptate reprezintă cheia în scăderea incidenței leziunilor de grad înalt și a cancerului invaziv. În țările în care se desfășoară programe de screening, acestea și-au dovedit eficiența, incidența cancerului de col uterin scăzând foarte mult. HPV (human papilloma virus) este considerat cauza majoră a cancerului de col uterin. Transformarea celulelor agresate viral în celule displazice și apoi în celule canceroase necesită însă și existența altor factori. Infecția cu HPV care este considerată o boală cu transmitere sexuală, apare mai frecvent în legătură cu: debutul precoce al vieții sexuale, promiscuitate, parteneri multipli, fumat. Există categorii cu risc de a dezvolta o leziune displazică și chiar cancer invaziv în urma infecției HPV, respectiv paciente cu nivel scăzut al imunității care nu pot elimina virusul sau localiza leziunile determinate de acesta. În același timp este important și genotipul viral, deoarece există peste 100 tulpini de HPV, dintre care aproximativ 40 au tropism pentru tractul genital inferior.

Pe criterii epidemiologice acestea au fost clasificate în:

- **cu risc scăzut (*low-risk*):** 6, 11, 40, 42, 43, 44, 54, 61, 70, 81 care rareori determină leziuni care progresează spre leziuni de grad înalt și cancer;
- probabil cu risc crescut: 26, 53, 66;
- **cu risc crescut (*high-risk*):** 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 și 82.

Cea mai eficientă și larg utilizată metodă, folosită în multe țări în cadrul programelor de screening pentru cancerul colului uterin, care se poate realiza relativ cu ușurință asupra unei populații largi, este reprezentată de citologia cervico-vaginală.

Evaluarea citopatologica este o tehnica sigura, eficienta si bine stabilita pentru diagnosticul multor afectiuni. Cel mai mare success in citologie este abilitatea acesteia de a reduce morbiditatea cancerului cervical prin screening populational. Unul din rolurile citologiei este acela de depistare precoce a displaziei sau celulelor canceroase preinvazive. Odata ce aceste celule sunt detectate, pacienta poate fi programata la o examinare prin biopsie si urmat de tratamentul chirurgical corespunzator. In consecinta progresia cancerului poate fi stopata intr-un stadiu timpuriu. În general, Pap testul se recomandă a fi efectuat periodic la intervale de 3 ani între 21 și 65 ani sau la 3 ani de la începutul vieții sexuale, deoarece se consideră că majoritatea infecțiilor cu HPV sunt contactate în primii ani de la debutul vieții sexuale (18-25 ani).

Diagnosticul clasic citologic este bazat pe observarea microscopica a celulelor specializate si interpretarea calitativa folosind criterii descriptive, care pot fi inconsistente din cauza variabilitatii subiective a diferitilor observatori. Pentru a scadea rata fals negativelor in screening au fost introduce tehnologii avansate in prelevare, pregatirea preparatului si controlul de calitate al screeningului. Dispozitivele comerciale care folosesc aceste tehnologii pot fi impartite in functie de abordarea lor in urmatoarele categorii:

1. Pregatirea lamelor imbunatatita pentru a reduce eroarea de prelevare a probelor si anume preparatele bazate pe strat subtire (ThinPrep, SurePath, Tripath)
2. Reducerea erorilor de screening ca in sistemele de autoscreening (ThinPrep Imaging System, CyTyc, FocalPoint System, Tripath Imaging)
3. Control de calitate imbunatatit in laborator ca de ex. in Papnet

Totuși, toate aceste dispozitive nu ofera o “diagnoza” obiectiva prin punerea la dispozitie a variabilelor calculabile care pot sa elimine erorile de interpretare sau discrepante inter-observator. In plus, nu pot fi aplicabile la un laborator general de citologie din cauza costurilor destul de mari si a golurilor tehnice si lingvistice. Din aceasta cauza, fara avea o unealta cantitativa, laboratoarele de citologie au in continuare o problema din cauza observatorilor subiectivi. Divergenta in diagnostic cauzata de observarea vizuala ramane in continuare o problema.

I . Baza de date initiala cu imagini propriu-zise si baza de date adnotata de specialisti histopatologi

In aceasta prima etapa de demarare a proiectului se urmareste obtinerea unei baze de date initiala cu imagini, necesara pentru inceperea studiului tehnic propus. Dincolo de costurile ceva mai mari, inca de la inceput, toate imaginile au fost achizitionate de la probe la care s-a efectuat citologie lichida si nu cea conventionala. Avantajele sunt enumerate mai jos, dar cel mai important aspect este acela ca aceasta tehnica este pretabila la automatizari, iar calitatea este ceva mai constanta si nu tine de subiectivitatea operatorului. Tehnicile de analiza de imagine computerizate folosite pentru asistenta in diagnosticul celulelor anormale sau a tumorilor din citopatologie si histopatologie pot sa evalueze destul de obiectiv morfologia nucleara. Metodele cantitative de estimare a specimenelor citologice au aparut in urma cu 30 de ani si continua sa se dezvolte. In orice design nou trebuie luate in considerare aspectele de accesibilitate, cost, eficienta, mentenanta tehnica si comunicare lingvistica. Datorita revolutiei si evolutiei noilor tehnologii, a puterii marite de calcul, scaderea costului partilor hardware si software, prevalenta internetului, sunt dezvoltate din ce in ce mai multe sisteme care folosesc algoritmi specifici de analiza de imagine. Aceasta abordare promite sa rezolve limita analizei subiective in mod special in domeniile bioinformaticii, biologiei si medicinei.

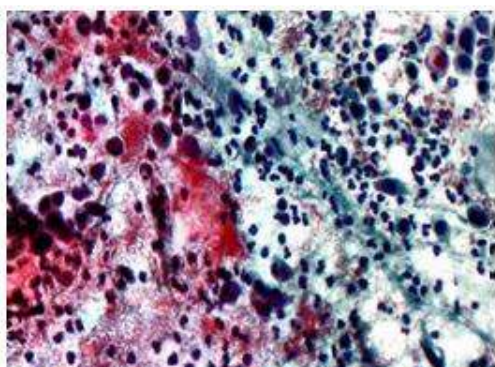


Fig.4: Frotiu convențional

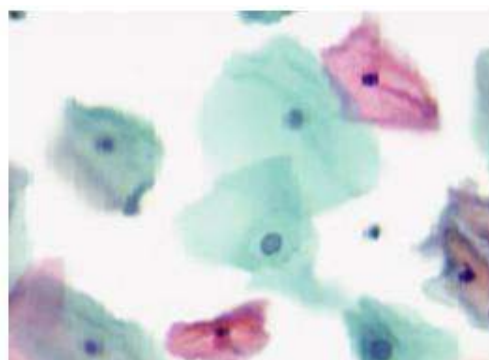


Fig.5: LBC

Avantajele LBC (Liquid Base Cytology – citologie în mediu lichid):

- recoltare facilă și rapidă, fără riscul distrugerii sau deteriorării celulelor;
- 100% din celulele epiteliale recoltate sunt transferate la laborator;
- fixare mai eficientă în mediul de transport;

- înlăturarea majorității factorilor de obscurare și realizarea unui frotiu alcătuit dintr-un singur strat de celule uniform distribuit, concentrat într-o zonă mică, facilitând interpretarea și îndeplinirea condițiilor de acceptabilitate;
- scăderea numărului de probe ce trebuie repetate în urma raportării ca fiind nesatisfăcătoare;
- compatibil cu sistemul automat de îmbogățire celulară, colorare și interpretare asistată de calculator.

Limite și interferențe:

- factori care țin de *calitatea tehnică*: nerespectarea condițiilor privitoare la pregătirea pacientelor sau la momentul recoltării. Ex: recoltare în perioada menstruală, folosire de lubrifiante intravaginale înaintea/în timpul recoltării etc.
- factori fiziologici sau patologici care țin de *paciente*: sarcină, lehoz, postpartum (interval de 6-8 săptămâni), atrofie (se poate recomanda terapie trofică locală cu estrogeni), radioterapie.
- *celularitatea scăzută* a frotiului poate limita posibilitatea elaborării unui diagnostic (criteriul acceptabilității). Totuși, preparatele se vor examina pentru a determina cauzele probabile ale acestei deficiențe, care în anumite condiții poate fi normală, sau pentru prezența celulelor atipice, caz în care se precizează în buletinul de analiză că preparatul are celularitate la limită sau scăzută. LBC și sistemul de „îmbogățire celulară” ajută la îndeplinirea acestui criteriu, dar există un minim necesar de 5000 de celule.
- *absența celulelor endocervicale sau metaplaziate* semnifică eșantionarea inadecvată a zonei endocervicale, iar aceasta se specifică în buletinul de raportare.
- *factorii de obscurare* pot interfera cu interpretarea atunci când afectează >75% (nesatisfăcător) sau 50-75% dintre celule (satisfăcător, dar parțial obscurat). Cea mai mare parte a acestor inconveniente sunt înlăturate prin recoltarea în mediu lichid³.
- calitatea celulelor importante pentru diagnostic poate fi influențată de manevre medico-chirurgicale locale practicate anterior recoltării: examinare manuală, aplicație lugol/acid acetic, recoltări anterioare, electroresecții etc.
- Pap testul reprezintă un instrument foarte util de screening folosit în depistarea precoce sau în stadii incipiente ale cancerului cervical, dar trebuie urmat întotdeauna de alte investigații înainte de a se lua o decizie terapeutică. Pap testul triază eficient pacientele care vor efectua investigații suplimentare (colposcopie) în vederea stabilirii de către medicul clinician a atitudinii terapeutice.

Criteriile traditionale citologice de diferentiere a celulelor displastice de cele normale sunt bazate pe modificarile raportului nucleu citoplasma, marimea nucleului, iregularitate in forma nucleului si a membranelor nucleare, densitate si granularitatea cromatinei nucleare. Toate aceste criterii sunt descriptive si sunt destul de subiective. Prin contrast, la morfometria computationala, criteriul subiectiv este inlocuit cu variabile comparative, calculabile si cuantificabile cantitativ.

Sistemul de analiză computerizată scanează toate frotiurile în mediu lichid sau convenționale și ajută la îmbunătățirea calității interpretării lor prin direcționarea atenției patologului asupra câmpurilor microscopice care, cel mai probabil, conțin anomalii, prin clasificarea frotiurilor conform unui sistem de cuantificare a riscului bazat pe măsurarea a sute de parametri celulari.

În general, există trei tipuri de celule cervicale scuamoase: celule superficiale, intermediare și parabazale care sunt observate în preparatele de Papanicolaou.

Prin acest sistem se realizează o detecție crescută a cazurilor HSIL-pozitive atât pe LBC, cât și pe preparatul convențional, precum și îmbunătățirea detecției LSIL pe cele convenționale.

Sistemul poate reduce semnificativ timpul de examinare și, implicit, timpul în care este elaborat diagnosticul final, fără a scădea rata de detecție a leziunilor importante clinic.

I.1. Baza de date cu imagini histopatologice

Prezentul proiect dorește să pună la punct o unealtă automată, reproductibilă și realistă pentru a facilita interpretarea preparatelor de citologie lichidă. Etapele parcurse până la obținerea imaginilor adnotate sunt următoarele:

Pregătire pacientă - cu 24-48 ore înainte de recoltarea probei trebuie să fie evitate: raporturile sexuale, lavajul vaginal, alte tratamente intravaginale (geluri, creme, contraceptive, dezinfectante, lubrifianti) sau alte manevre intravaginale (tampoane intravaginale, explorare vaginală).

- Se recomandă recoltarea în afara perioadei menstruale, în perioada de mijloc a ciclului menstrual, iar în cazul infecțiilor, după tratarea acestora.

Specimen recoltat - celule ale zonei exocervicale, endocervicale și de tranziție dintre acestea (zona de transformare). Zona de transformare reprezintă sediul cel mai frecvent al

leziunilor preneoplazice (turn-over crescut, expusă microtraumatismelor, sensibilitate crescută la acțiunea HPV), acolo unde epiteliul scuamos al exocolului se continuă cu cel de tip glandular de la nivel endocervical prin intermediul unei zone de metaplazie scuamoasă.

Recoltare - se efectuează cu ajutorul periutei cervicale cu cap detașabil. Pentru recoltarea în mediu lichid sunt recomandate dispozitivele din plastic. După recoltare, partea detașabilă a instrumentului este introdusă în flaconul cu mediu lichid, fără a se efectua alte manevre asupra suprafeței care conține materialul celular.



Fig.1: Periută Cervex-Brush¹⁰



Fig. 2: Flacon fixare și transport¹⁰



Fig.3: Detașare capăt recoltare

Instrucțiuni de recoltare – După vizualizarea colului uterin se inserează partea centrală a periutei în canalul endocervical astfel încât marginile mai scurte să vină în contact cu exocervixul. De obicei, nu se recomandă ștergerea colului uterin înainte de recoltare, dar la nevoie se poate tampona ușor cu un tifon. Proba astfel recoltată trebuie imediat fixată pentru a se evita deteriorarea probei. În acest scop, se detașează capătul periutei care conține materialul recoltat și se introduce în recipientul cu mediu lichid. Se acoperă flaconul cu capacul, se înșurubează pentru a nu se varsa lichidul, se etichetează imediat

(sau înainte de recoltarea probei) și se transportă la laborator împreună cu fișa de însoțire a pacientei în vederea prelucrării.

Este foarte important ca, pe lângă corectitudinea recoltării, proba să fie însoțită și de datele clinice cât mai complete ale pacientei: vârsta, data ultimei menstruații, data recoltării, antecedente medicale, tratamente, metode de contracepție, rezultate anterioare ale Pap testului etc.

Metodă - citologie în mediu lichid, colorație Papanicolaou.

Toate analizele și experimentele au fost făcute într-o configurație de laborator de citologie clasică. Probele au fost prelevate de la paciente care au fost îndrumate de către medicii ginecologi pentru controale de rutină periodice. La cele mai multe s-a făcut în paralel și controlul infecției cu virusul HPV.

Lamele au fost analizate de către anatomopatologi după care au fost preluate imagini pe un microscop Olympus C21 cu camera Olympus C7070 la o rezoluție de 7 Mpixeli. Pentru fiecare pacientă s-au achiziționat cel puțin 30 de câmpuri de microscopie la o mărire de 20X. Formatul imaginilor a fost JPEG pentru a micșora mărimea fișierelor.

I.2. Adnotarea imaginilor de către anatomopatologi

Sau preluat în jur de 900 de imagini de la lame cu preparate PAP normale. Lamele au provenit de la 30 de paciente fără complicații majore, în vederea furnizării spre analiză de imagine a lotului martor. Fiecare imagine a fost adnotată și datele au fost trecute într-un tabel excel cu următoarele categorii în capul de tabel:

1. Nr. Imagine
2. Nr. Caz
3. Căi de frotiu
4. Celule endocervicale
5. Hematii
6. Celule Metaplazice
7. Modificări celulare
8. Modificări Celulare HPV
9. Herpes
10. Bacterii A
11. Vaginoza

- 12. Cocobacili
- 13. Tichomonas
- 14. Candida
- 15. Rezultat

În fiecare câmp s-au introdus valorile „1” pentru prezenta sau „0” pentru lipsa. La câmpul calitate frotiu „1” înseamnă imagine pretabilă la analiză și „0” calitate precară sau necorespunzătoare.

Rezultate, *elaborarea diagnosticului*:

În urma prelucrării și examinării probei, rezultatul se elaborează conform Sistemului de raportare a citologiei colului uterin Bethesda 2001.

Imaginile preluate de la cazurile cu diagnostic stabilit au fost folosite pentru stabilirea bazei de date incipente (care este în continuă creștere).

Termeni utilizați:

- acceptabilitatea specimenului;
- modificări non-neoplazice – NILM;
- anomalii ale celulelor epiteliale scuamoase: ASC-US, ASC-H, LSIL, HSIL, SCC;
- anomalii ale celulelor glandulare-AGC, AIS, adenocarcinom.

1. ACCEPTABILITATEA SPECIMENULUI: criteriile necesare a fi îndeplinite de un preparat pentru a fi raportat ca fiind *satisfăcător* pentru interpretare sunt:

- **de celularitate minimă:** chiar dacă metoda LBC înlătură cea mai mare parte dintre inconvenientele frotiurilor convenționale, asigurând ajungerea în laborator a 100% din materialul recoltat și selectarea masei celulare importante pentru diagnostic, există un minim necesar de celule scuamoase bine conservate și ușor de vizualizat pentru elaborarea diagnosticului (5000).
- **prezența celulelor endocervicale sau din zona de transformare:** indică recoltarea corespunzătoare a frotiului cu eșantionarea zonei de transformare (minim 10 celule endocervicale sau metaplaziate bine conservate, izolate sau grupate). Precizarea prezenței/ absenței celulelor endocervicale/metaplaziate nu este necesară în leziunile de grad înalt sau invazive.
- **factori de obscurare:** sânge, mucus etc. Se consideră că un frotiu este *nesatisfăcător* atunci când >75% dintre celulele scuamoase sunt acoperite și *parțial obscurat* dar

satisfăcător pentru interpretare dacă 50-75 % dintre celule sunt obscurate, cu condiția ca pe frotiu să nu fi fost identificate celule atipice. LBC are avantajul că înlătură cei mai mulți dintre acești factori.

*ASCCP 2007 recomandă în caz de teste nesatisfăcătoare repetarea la 2-4 luni, repetare la 12 luni pentru cele cărora le lipsește componenta endocervicală sau parțial obscurate⁷.

2. NILM - frotiu negativ pentru leziuni intraepiteliale sau maligne;

- **microorganisme:** paraziți (*Trichomonas vaginalis*), organisme fungice compatibile cu *Candida*, modificări ale florei vaginale compatibile cu diagnosticul de vaginoză, bacterii sugestive pentru genul *Actinomyces*, modificări celulare sugestive pentru infecția cu Virus Herpes Simplex (HSV);
- **alte modificări non-neoplazice:** modificări celulare reactive asociate cu: inflamația, iradierea, dispozitive contraceptive intrauterine; prezența de celule glandulare (status posthisterectomie), atrofie.

3. ALTELE: *celule endometriale la femei >40 ani.*

4. ANOMALII ALE CELULELOR EPITELIALE:

- **SCUAMOASE:**
- **Celule scuamoase atipice:** - cu semnificație nedeterminată (**ASC-US**);
- nu se poate exclude HSIL (**ASC-H**).
- **Leziuni scuamoase intraepiteliale de grad scăzut (LSIL)** - includ modificări induse de HPV; displazia ușoară; CIN1.
- **Leziuni scuamoase intraepiteliale de grad înalt (HSIL)** - includ leziunile: CIN2 (displazia moderată) și CIN3 (displazia severă și carcinomul in situ).
- **Carcinom cu celule scuamoase.**
- **GLANDULARE:**
- Celule atipice (endocervicale, endometriale, glandulare) cu semnificație neprecizată (**AGC-NOS**).
- celule atipice (endocervicale, glandulare) sugestive pentru neoplazie (**FN**).
- adenocarcinom endocervical in situ (**AIS**).
- adenocarcinom: endocervical, endometrial, extrauterin, cu semnificație neprecizată (**NOS**).

5. Alte leziuni maligne.

Interpretarea rezultatelor, recomandări:

Rezultatul Pap testului trebuie interpretat prin corelarea cu datele clinice ale pacientelor precum și cu alte investigații efectuate. Astfel, în funcție de suspiciunea ridicată în urma Pap testului, prin integrarea în contextul clinic de către medicul curant, se pot formula recomandările privitoare la conduita terapeutică și diagnostică ce trebuie urmată. Asociațiile profesionale medicale din domeniu au elaborat o serie de recomandări asupra atitudinii de urmat în funcție de rezultatele Pap testului.

- **NILM** - semnifică absența leziunilor intraepiteliale sau a malignității. Această categorie este folosită și pentru a raporta prezența colonizării cu microorganisme, dar și a leziunilor reactive sau reparatorii ale celulelor ca răspuns la infecții, iradiere, DIU etc. Raportarea celulelor glandulare benigne posthisterectomie poate semnifica: adenoză, prolaps salpingian, metaplazie etc. Tot în categoria NILM sunt incluse și modificările induse de atrofie.
- **Celulele endometriale la femei cu vârste peste 40 ani.** Acestea pot sugera riscul unei patologii endometriale care pentru femeile în premenopauză este mic (în general patologie benignă: polipi, hiperplazie simplă), în timp ce la femeile în postmenopauză ele pot indica necesitatea investigării endometrului prin corelarea cu semnele și simptomele clinice, terapii hormonale etc. Prezența celulelor endometriale este normală pe speciimenele menstruale sau din timpul fazei proliferative a ciclului menstrual.
- **CELULE SCUAMOASE ATIPICE:**
 - ASC-US:** - semnifică prezența de celule epiteliale atipice a căror semnificație nu poate fi precizată. Modificările celulelor epiteliale sunt mai importante decât cele din procesele inflamatorii, dar nu suficiente pentru interpretarea ca leziune intraepitelială.
 - prezintă un risc pentru o leziune de grad înalt de 10-20%.
 - ASC-H:** - sugerează prezența unor celule epiteliale atipice sugestive pentru leziuni intraepiteliale de grad înalt, dar care nu întrunesc pe deplin criteriile cantitative sau calitative pentru calificarea ca HSIL.
 - 50% din citologiile cu ASC-H ascund riscul unei leziuni de grad înalt;
 - LSIL:** - sugerează modificări ale celulelor epiteliale scuamoase compatibile cu o leziune intraepitelială de grad scăzut. Termenul de leziune intraepitelială se referă la faptul că modificările celulare sunt strict localizate la nivelul epiteliului de acoperire al colului uterin, fără a depăși membrana bazală, deci fără potențial invaziv. Într-o leziune

de grad scăzut modificările atipice sunt prezente doar la unele dintre celulele scuamoase.

- în această categorie sunt încadrate modificările celulare induse de HPV și CIN1 (displazia ușoară);

HSIL: - sugerează prezența unei leziuni intraepiteliale de grad înalt în care atipiile celulare sunt mai mari și afectează o proporție mai mare din celulele epiteliale, dar sunt încă limitate la epiteliul de acoperire al colului uterin fără a depăși membrana bazală.

- apare ca urmare a persistenței și evoluției unei leziuni de etiologie HPV.

- include modificări CIN2 (displazia moderată) și CIN3 (displazia severă și CIS).

SCC: - indică prezența unei leziuni neoplazice invazive care a depășit stadiul de leziune intraepitelială prin extensia dincolo de membrana bazală.

- **CELULE GLANDULARE ATIPICE:**

AGC: - această categorie reunește modificările atipice ale celulelor epiteliului glandular superioare celor din leziunile reactive sau reparatorii, dar insuficiente pentru neoplazie.

- dacă este posibilă identificarea originii, atunci se precizează în raport, în caz contrar sunt încadrate în categoria **NOS**.

- dacă modificările celulare sunt sugestive pentru prezența unei leziuni neoplazice, se precizează în raport ca fiind *sugestive* pentru neoplazie. Specificația NOS este folosită atunci când nu este posibilă precizarea potențialului neoplazic al celulelor glandulare atipice.

AIS: - semnifică prezența unei leziuni neoplazice limitată la epiteliul glandular endocervical fără a depăși membrana bazală.

Adenocarcinomul endocervical, endometrial sau extrauterin (ovarian, salpingian): este o leziune care a depășit epiteliul de suprafață și are capacitate invazivă.

- **Alte neoplasme:** în această categorie sunt incluse tumorile primare rare ale colului uterin (carcinomul cu celule mici, carcinosarcomul, tumori germinative, limfoame etc.), precum și cele secundare (endometru, rect, vezică urinară) sau de origine metastatică (sân, melanom etc.).

Abrevieri

AGC – celule glandulare atipice

AIS – adenocarcinom in situ

ASC – celule scuamoase atipice

ASC-US – celule scuamoase atipice cu semnificație nedeterminată
ASC-H – celule scuamoase atipice fără a se putea exclude o leziune de grad înalt
ASCCP – American Society of Colposcopy and Cervical Pathology
CIN – neoplazie cervicală intraepitelială
CIS – carcinom in situ
HPV – virusul papilloma uman
HSIL – leziune scuamoasă intraepitelială de grad înalt
LBC – Liquid Base Cytology – citologie în mediu lichid
LSIL – leziune scuamoasă intraepitelială de grad scăzut
NILM – negativ pentru leziuni maligne intraepiteliale
NOS – (no other specifications) fără alte specificații
SCC – carcinomul scuamocelular
TBS – The Bethesda System

Bibliografie

1. Diane Solomon, Ritu Nayar. Sistemul Bethesda de raportare a citologiei colului uterin, Ed. Medicală Calisto, 2009.
2. M. Arbyn, A. Anttila, J. Jordan, G. Ronco, U. Schenck, N. Segnan, H. Wiener, A. Herbert & L. von Karsa. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition—Summary Document, In Annals of Oncology 21, 2010: 448–458
3. M. Arbyn, A. Herbert, U. Schenck, P. Nieminen, J. Jordan, E. Mcgoogan, J. Patnick, C. Bergeron, J-J. Baldauf, P. Klinkhamer, J. Bulten and P. Martin-Hirsch. European guidelines for quality assurance in cervical cancer screening: recommendations for collecting samples for conventional and liquid-based cytology. In Cytopathology 2007, 18, 133– 139.
4. Marluce Bibbo, David C. Wilbur, Comprehensive Cytopathology, Third Edition, Saunders Elsevier 2008.
5. Richard Mac DeMay, The Pap Test, ASCP Press, Chicago, 2005.

Raport de cercetare UPB

La UPB activitatea de cercetare a început prin studierea literaturii publicate în domeniu. S-a observat ca unul din factorii principali care stau la baza interpretării frotiurilor Papanicolaou de către medicii anatomopatologi este analiza nucleelor celulelor [1, 2, 3], drept pentru care am demarat studiul acestei probleme. Un factor foarte important pentru succesul unei operațiuni de segmentare este calitatea imaginii ce trebuie segmentată, drept pentru care am investigat de asemenea și posibilitatea îmbunătățirii imaginilor obținute prin scanarea frotiurilor Papanicolaou. În continuare vom detalia principalele rezultate obținute pentru îmbunătățirea imaginilor și pentru segmentarea automată a nucleelor celulelor.

1. Segmentarea automată a nucleelor în imagini histopatologice ale colului uterin

Segmentarea nucleelor celulelor în imagini anatomopatologice a făcut obiectul a numeroase cercetări până în prezent [1-7]. În [1], este propusă o metodă de segmentare pe baza unui mecanism bazat pe logica fuzzy. În [3], autorii abordează problema printr-o binarizare bazată pe metoda graph-cut, urmată de o detecție de seed-uri printr-o metodă multiscală de filtrare LoG (Laplacian-of-Gaussian). Tot pe o detecție de seed-uri (obținute prin metode morfologice aplicate atât pe imaginea de intensitate, cât și pe cea de gradient) se bazează și metoda propusă în [2], doar că seed-urile sunt folosite ca punct de plecare pentru un algoritm de segmentare watershed cu markeri. Articolul [4] tratează special separarea celulelor adiacente, care, în mod normal, sunt detectate de o operație de segmentare ca fiind o singură celulă. În [6], în contextul detecției neuroblastomului, se propune o metodă de segmentare a nucleelor pe baza transformatei top-hat în combinație cu o segmentare cu histerezis.

Toate articolele studiate raportează scoruri de detecție foarte bune, dar performanțele lor depind extrem de mult de calitatea imaginilor de start. Cu alte cuvinte, în prezența unor imagini histopatologice în care contrastul între nucleee și celule scade, randamentul metodelor studiate începe să scadă (uneori dramatic). Scopul nostru este de a dezvolta un clasificator care să fie independent la variații de contrast ale imaginii sau la modul în care a fost preparată proba histopatologică în laborator, drept pentru care este necesară o operație de segmentare a nucleelor care să fie robustă la astfel de variații.

1.1 Metoda de segmentare propusă

Metoda de segmentare pe care am dezvoltat-o (și care până acum dă rezultate extrem de încurajatoare, deși nu este încă finalizată) și pe care o vom prezenta în acest raport este bazată pe metoda de segmentare *mean-shift* a lui Comaniciu și Meer [8]. Propusă inițial pentru segmentarea

imaginilor color, principiul de bază al metodei este de identificare a modurilor dintr-o histogramă (care poate fi multidimensională) printr-un algoritm de “ascensiune” din fiecare punct al histogramei respective către cel mai apropiat vârf al histogramei. Dacă asociem histograma respectivă cu un relief, algoritmul mean-shift identifică munții din relieful respectiv (un munte fiind asociat unui semn vârf). Algoritmul a fost folosit cu succes la segmentarea imaginilor color, prin identificarea modurilor din histograma color (2D sau 3D) a imaginilor respective.

Deși imaginile noastre histopatologice sunt color, totuși nu culoarea este relevantă pentru identificarea nucleelor, ci faptul că, întotdeauna, acestea sunt mai întunecate decât restul celulei (v. Figura 1).

În același timp, se poate observa că niciuna dintre tehnicile clasice de segmentare (prăguire pe niveluri de gri sau utilizare a gradientului pentru extragerea contururilor) nu are cum separa nucleele, având în vedere variația destul de mare a intensității luminoase a pixelilor aparținând nucleelor, precum și faptul că zona de frontieră între nucleee și restul celulelor este extrem de slab nuanțată (e blurată). Este evident că numai o tehnică de segmentare spațială, care ține cont în același timp atât de intensitate, cât și de poziția spațială a pixelilor poate da rezultate. O reprezentare a imaginilor care ține cont atât de poziția (i, j) a unui pixel, cât și de intensitatea lui $I(i, j)$ este de a reprezenta pixelul respectiv ca un punct într-un spațiu 3-dimensional având coordonatele $(i, j, I(i, j))$ (vezi Figura 2). Avantajul unei astfel de reprezentări a imaginii este că punctele alăturate care sunt de nivel de gri relativ distinct decât cele din vecinătatea lor vor forma un cluster de puncte separat, care poate fi identificat printr-un algoritm de tip mean-shift.

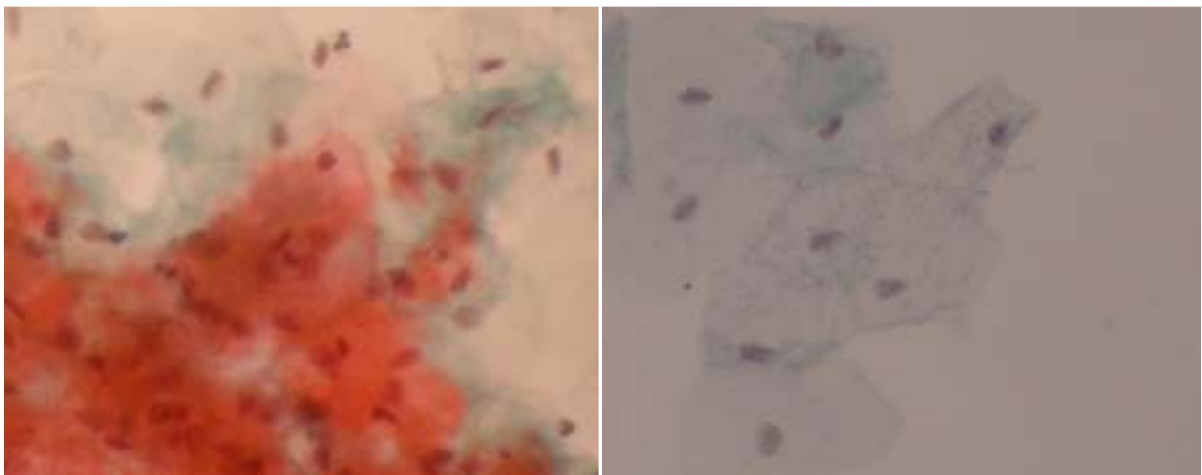


Figura 1. Exemple tipice de imagini histopatologice ale colului uterin. Se poate observa că nucleele celulelor diferă de interiorul acestora prin faptul că sunt mai întunecate, nu prin culoare.

Ideea de segmentare a nucleelor folosind tehnica mean-shift nu este, deci, de a folosi histograma color a imaginilor, ci de a transforma imaginile în imagini scalare (cu nivele de gri, în care un pixel

este reprezentat de o singură valoare) și de a aplica segmentarea mean-shift pe norul de puncte 3D echivalent imaginii respective (este de observat că tehnica mean-shift poate fi aplicată pe nori de puncte N-dimensionale, cu alte cuvinte, am putea păstra culoarea fiecărui pixel ca atare și caracteriza fiecare pixel ca fiind un punct într-un spațiu 5-dimensional $(i, j, R(i,j), G(i,j), B(i,j))$, dar asta ar implica o creștere foarte mare a timpului de calcul).

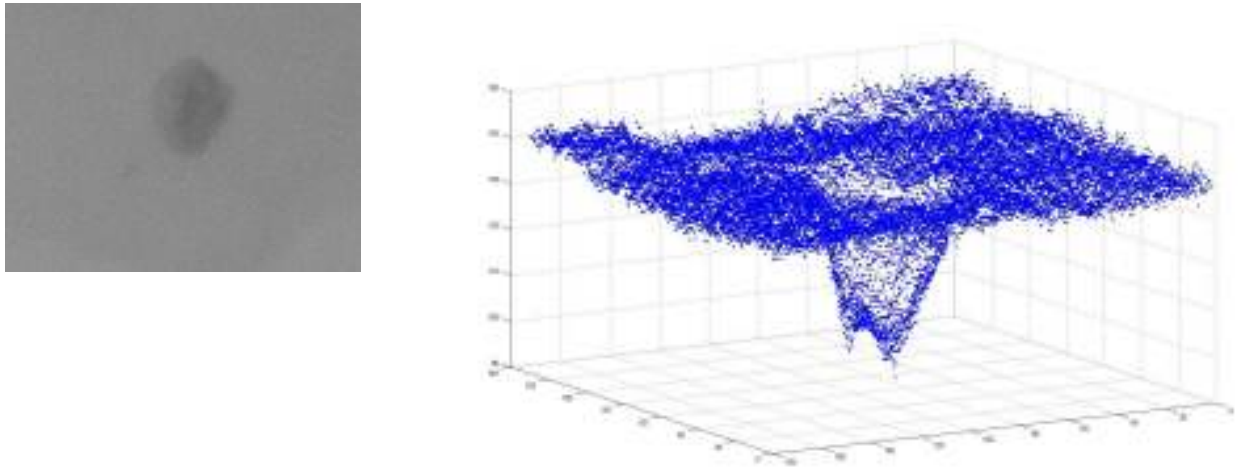


Figura 2. Reprezentarea unei imagini scalare sub formă de nor 3D de puncte

Tehnica implică calculul, pentru fiecare punct, a unei densități locale de puncte (prin convoluția norului de puncte cu un nucleu gaussian 3D), și “urcarea” pe gradientul maxim de densitate: alegerea vecinului 3D cu cea mai mare densitate locală, după care operațiunea se repetă, până când ajungem în vârf, adică în punctul caracterizat de o densitate locală de puncte maximă în vecinătatea sa. Toate punctele de pe calea parcursă, precum și toate punctele care (pe alte căi) ajung în același vârf local, vor fi etichetate ca aparținând acelui cluster (aceleiași regiuni). De remarcat faptul că există un singur parametru de reglaj al metodei, și anume lățimea σ a gaussienei cu care se calculează densitatea locală de puncte. Figura 3 prezintă un rezultat tipic (brut, adică fără nicio prelucrare ulterioară a regiunilor) al segmentării propuse.

După cum se poate observa, majoritatea nucleelor prezente în imagine sunt separate de fond (și asta în condiții extrem de grele: contrast slab între nucleu și fond cu frontieră blurată), separare esențială pentru segmentarea finală a nucleelor.

Se observă totodată că sunt numeroase artefacte în imaginea segmentată brută, printre care cel mai important este că, în majoritatea cazurilor, nucleele sunt suprasedgmentate în mai multe regiuni. Această problemă trebuie rezolvată printr-un procedeu de unificare de regiuni (*region merging*) ce trebuie aplicat post-segmentare. O altă problemă care trebuie rezolvată este de a decide dacă o

regiune este sau nu nucleu. În continuare acestui raport vom prezenta soluțiile identificate pentru cele două probleme.

În ceea ce privește unificarea regiunilor, într-o fază inițială, aplicăm o deschidere morfologică binară (operație ce constă în aplicarea unei erodări, urmată de o dilatare) fiecărui segment. Această operațiune va elimina numeroasele regiuni de dimensiuni de până la 5 pixeli, pe care le va realoca uneia din regiunile mai mari învecinate (în condițiile în care două regiuni mari, în urma dilatării, vor tinde să înglobeze un pixel care nu mai aparține niciunei regiuni, atunci acesta va fi alocat regiunii mari de care este cel mai apropiat din punct de vedere al intensității).

Apoi, pentru regiunile rămase, se face o analiză a posibilităților de unificare cu regiunile vecine. În acest sens, se calculează un grad de adiacență (A) între două regiuni distincte, bazat pe înălțimea (h) și înălțimea (w) a celor trei dreptunghuri de încadrare (câte unul pentru fiecare regiune, și unul pentru regiunea formată din unificarea spațială a celor două regiuni):

$$A = \max\left(1 - \frac{w_1 + w_2}{w_u}, 1 - \frac{h_1 + h_2}{h_u}\right).$$

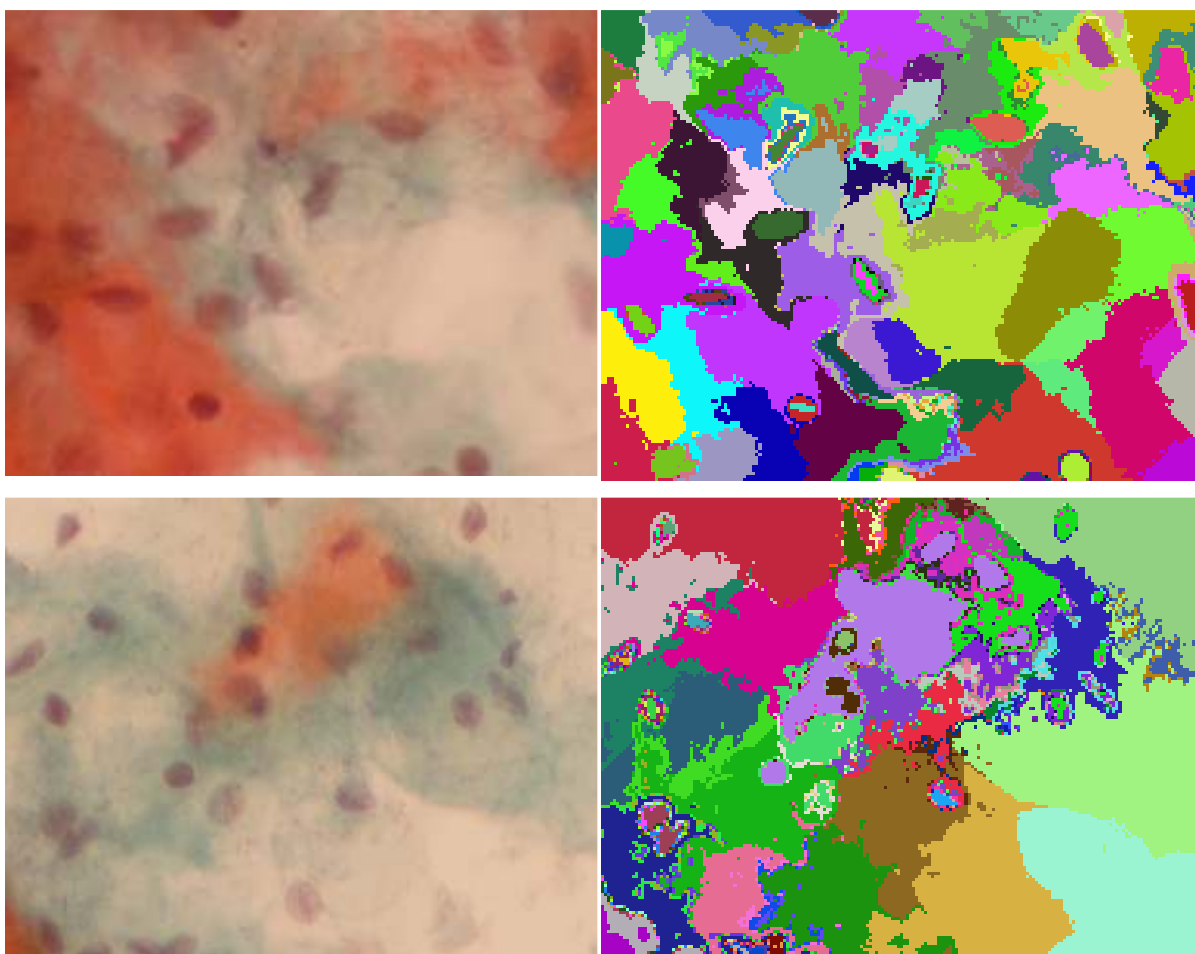


Figura 3. Exemple tipice de segmentare mean shift pe norul de puncte 3D. Imaginea segmentată

(coloana din dreapta) este prezentată în pseudoculori (fiecare segment este afișat cu o culoare aleatoare).

Evident, dacă dreptunghiurile de încadrare ale cele două regiuni se suprapun, avem simultan $w_1+w_2>w_u$ și $h_1+h_2>h_u$, deci, implicit, un grad de adiacență negativ; pe măsură ce distanța spațială între regiuni crește, crește și gradul de adiacență A . Se formează toate grupurile (din două, trei, sau mai multe) regiuni cu grad (cumulat) de adiacență mic și se testează în vederea unificării (trebuie menționat că la această operațiune nu participă decât regiunile cu dimensiune relativ mică, pentru a reduce timpul de calcul alocat acestei etape). Criteriul pentru unificare este diferența între intensitățile medii ale regiunilor, precum și media modului gradientului de-a lungul frontierelor între regiuni.

În ceea ce privește decizia nucleu/non-nucleu pentru fiecare regiune rămasă în harta de regiuni, luăm în calcul următorii parametri:

- factorul de aspect al regiunii (am încercat mai multe moduri de calcul al acestuia, fie ca raport între perimetru și arie, fie ca raport al valorilor proprii ale matricii de covariație spațială a coordonatelor (i,j) ale pixelilor din regiune);
- dimensiunea regiunii
- contrastul local, calculat ca:

$$C = \frac{\bar{I}_{ext} - \bar{I}_{reg}}{\bar{I}_{ext} + \bar{I}_{reg}}$$

unde, $\bar{I}_{ext}$ este media intensității regiunilor ce înconjură segmentul iar $\bar{I}_{reg}$ este valoarea medie a regiunii în cauză.

Regiunile care sunt suficient de rotunde, au o dimensiune rezonabilă, și care au un contrast local mai mare decât un prag sunt declarate nuclee.

1.2 Rezultate

Rezultatele obținute de metoda de segmentare propusă sunt încurajatoare. Pe un set de 30 de imagini pe care le-am marcat manual (multe dintre ele cu nuclee dificil de identificat chiar și de operatorul uman), au fost identificate 82% din nucleele prezente, la o rată de fals pozitive (regiuni identificate greșit ca fiind nuclee de celule) de 3%. Dacă relaxăm pragurile astfel încât să crească rata de detecție începe să crească destul de mult rata de fals pozitive (de exemplu la o rată de detecție de 85%, se înregistrează în prezent o rată de fals pozitive de 10%). Un exemplu este prezentat în Figura 4.

Principalele cauze ale nereușitelor sunt:

- contrastul extrem de scăzut între nucleu și restul celulei, astfel încât regiunea, deși identificată, nu trece de pragul de contrast impus;
- nesepararea (de către segmentarea mean-shift) de către nucleu de restul celulei, tot din motive de contrast extrem de scăzut.
- faptul că, în unele cazuri, nucleeele efectiv nu sunt separate spațial, ceea ce conduce la o regiune detectată care nu corespunde din punct de vedere geometric (e fie prea mare, fie are o formă necorespunzătoare).

De asemenea, o altă problemă importantă este complexitatea computațională. În prezent, din lipsă de memorie, metoda nu poate fi rulată pe întreaga imagine disponibilă, de 6 milioane de pixeli, pe un PC dotat standard (pentru anul 2014) întrucât numai spațiul de memorie necesar stocării hărții 3D a densității ar fi de peste 2 GB.



Figura 4. Exemplu de segmentare a nucleelor pentru imaginea din Figura 3 (linia a doua, coloana 1)

2. Îmbunătățirea imaginilor

Activitatea legată de îmbunătățirea imaginilor a fost legată strict de operațiunea de segmentare descrisă în secțiunea precedentă.

În primul rând, una din problemele de care ne-am lovit a fost cum să transformăm optim imaginea color disponibilă într-una scalară, pentru a putea aplica segmentarea pe norul 3D de puncte rezultat. Dintre toate testele efectuate, a reieșit că cel mai bine sunt separate nucleeele de restul celulei în componenta de verde a tripletului RGB (mai bine decât orice combinație liniară între cele trei

componente de culoare). Drept pentru care am decis ca operațiunea de segmentare să se efectueze numai pe componenta de verde a culorii.

O a doua problemă pe care am studiat-o a fost îmbunătățirea rezultatelor segmentării prin pre-procesarea imaginii. Dintre tehnicile testate, menționăm:

- îmbunătățirea contrastului prin egalizare de histogramă
- îmbunătățirea contrastului prin egalizare de histogramă la nivel local și global
- îmbunătățirea contrastului prin operatori de accentuare (*sharpening*)
- netezirea regiunilor prin aplicarea de filtre liniare
- netezirea regiunilor prin aplicarea de filtre ne-liniare care împiedica blurring-ul de frontiere (filtru bilateral, filtru LLSMEE etc.).

Pentru fiecare metodă în parte, am testat rezultatele segmentării pe baza de date marcată, cu și fără metoda respectivă. Deși în ceea ce privește aspectul propriu-zis al imaginilor (evaluate de ochiul uman) impresia era clară de îmbunătățire a calității (mai ales pentru creșterea liniară a contrastului) cifrele (seci) au dezmințit impresia inițială, în sensul că s-a constatat că performanțele metodei de segmentare propuse nu au fost îmbunătățite de niciuna dintre metodele testate.

3. Continuarea cercetării

Pentru perioada imediat următoare, intenționăm să rafinăm metoda de segmentare propusă, o direcție fiind îmbunătățirea rezultatelor, iar cealaltă fiind reducerea cerințelor computaționale.

În paralel, vom începe studiul clasificatorilor ce pot fi folosiți pentru a decide asupra zonelor cu potențial malign din imaginile disponibile.

Nu în ultimul rând, intenționăm ca metoda dezvoltată, din momentul în care va fi finalizată, să o comunicăm prin intermediul unui articol la o conferință internațională.

4. Referințe bibliografice

- [1] Begelrnan, G., et al. "Cell nuclei segmentation using fuzzy logic engine." 2004 International Conference on Image Processing, 2004. ICIIP'04. Vol. 5. IEEE, 2004.
- [2] Wählby, Carolina, et al. "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections." *Journal of Microscopy* 215.1 (2004): 67-76.
- [3] Al-Kofahi, Yousef, et al. "Improved automatic detection and segmentation of cell nuclei in histopathology images." *IEEE Transactions on Biomedical Engineering*, 57.4 (2010): 841-852.
- [4] Li, Gang, et al. "Segmentation of touching cell nuclei using gradient flow tracking." *Journal of Microscopy* 231.1 (2008): 47-58.

- [5] Yang, Xiaodong, Houqiang Li, and Xiaobo Zhou. "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy." IEEE Transactions on Circuits and Systems I: Regular Papers, 53.11 (2006): 2405-2414.
- [6] Gurcan, Metin N., et al. "Image analysis for neuroblastoma classification: Segmentation of cell nuclei." 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS'06. IEEE, 2006: 4844-4847
- [7] Meijering, Erik. "Cell segmentation: 50 years down the road [life sciences]." IEEE Signal Processing Magazine, 29.5 (2012): 140-145.
- [8] Comaniciu, Dorin, and Peter Meer. "Mean shift: A robust approach toward feature space analysis." IEEE Transactions on Pattern Analysis and Machine Intelligence, 24.5 (2002): 603-619.