
Detekcija premalignih stanica

Nikolina Duranec, Ana Nedić, Borna Radoš, Fran Špigel, Prirodoslovno-matematički fakultet u Zagrebu

Rak vrata maternice zauzima četvrto mjesto po učestalosti smrtnog ishoda žena u svijetu. Procjenjuje se da je u 2018. godini od raka vrata maternice oboljelo 570 000 žena u svijetu te da je oko 311 000 žena umrlo od te bolesti. Visoka smrtnost raka vrata maternice može se smanjiti na globalnoj razini ranim otkrivanjem, redovitim odlascima na preglede i programima liječenja. Najvažniji rizični čimbenik raka vrata maternice u svijetu je infekcija humani papiloma virusom (HPV). U većini slučajeva HPV infekcija prolazi spontano, djelovanjem vlastitog imunološkog sustava, dok dugotrajnija infekcija može izazvati rak vrata maternice.

Ukoliko se otkrije u ranom stadiju razvoja, rak vrata maternice jedan je od najuspješnije liječivih vrsta karcinoma. Najzastupljenija metoda prevencije raka vrata maternice od pedesetih godina prošlog stoljeća postaje Papa-test (engl. pap smear test) koji analizom citologije vrata maternice otkriva predstadije i rane stadije raka vrata maternice. Nakon uzimanja staničnog uzorka za Papa-test, ginekolog ga nanosi na predmetno stakalce i šalje u citološki laboratorij gdje se dobiveni uzorak priprema na analizu. Stakalce se oboji metodom po Papanicolaou (engl. pap smear) koju je izumio otac cervikovaginalne citologije, George N. Papanicolaou. Boje se jezgra i citoplazma stanica. Specijalist kliničke citologije postavlja dijagnozu na temelju morfoloških osobina stanica u dobivenom uzorku. Karakteristike jezgre i citoplazme svake od stanica uspoređuju se s

karakteristikama normalnih stanica te se detektiraju abnormalne stanice.

Herlev

S Odjela za patologiju Sveučilišne bolnice Herlev u suradnji s Odjelom za automatizaciju Tehničkog sveučilišta u Danskoj prikupljeni su podatci od ukupno 917 slika od kojih svaka slika reprezentira jednu stanicu vrata maternice dobivenu Papa-testom. Na slici se nalaze jezgra i citoplazma stanice te prostor pozadine. Stanice su ručno razvrstali citotehničari i liječnici u sedam klasa. Obzirom da je s medicinskog gledišta najvažnija podjela na normalne i abnormalne stanice, upravo će ta binarna klasifikacija biti tema ovog rada.

Displastične su stanice one koje imaju veliki potencijal da postanu kancerogene te ih svrstavamo u abnormalnu klasu. Obzirom na njihov stupanj promjena, razlikujemo blagu displaziju, umjerenu displaziju i tešku displaziju (Mild Squamous non-keratinizing Dysplasia, Moderate Squamous non-keratinizing Dysplasia, Severe Squamous non-keratinizing Dysplasia). Abnormalnoj klasi pripada i takozvani karcinom in situ (Squamous cell carcinoma in Situ) kod kojeg postoji visoki rizik od pojave karcinoma.

Tri su tipa stanica koje pripadaju normalnoj klasi: kolumnarne, intermedijarne i superficijalne stanice (Superficial squamous epithelial, Intermediate squamous epithelial, Columnar epithelial). Podjela je vidljiva u Tablici 1.

Tablica 1: Podatkovni skup Herlev

klasa	broj stanica
ABNORMALNE STANICE	
blaga displazija	182
umjerena displazija	146
teška displazija	197
karcinom in situ	150
ukupan broj abnormalnih	675
NORMALNE STANICE	
kolumnarne	98
intermedijarne	70
superficialne	74
ukupan broj normalnih	242

Uz podatkovni skup Herlev se uz svaku originalnu sliku stanice nalazi i njena segmentirana slika kojom je napravljena podjela na staničnu jezgru i citoplazmu. Podjela na te dvije regije je bitna stavka koja citolozima pomaže pri označavanju je li stanica normalna ili kancerogena. Svjetloplavom bojom označena je jezgra, tamnoplavom citoplazma stanice, a crvenom bojom označen je prostor izvan stanice, odnosno pozadina. Dakle, ukupno je 917 slika originalnih stanica iz skupa podataka Herlev koje su izolirane iz uzoraka Papa-testa te je 917 segmentiranih, odnosno kontrastiranih slika. Primjeri originalnih i segmentiranih stanica dani su na Slici 1. Pretpostavljamo da će veličina jezgre (u odnosu na veličinu stanice) igrati veliku ulogu u detekciji premalignih stanica.

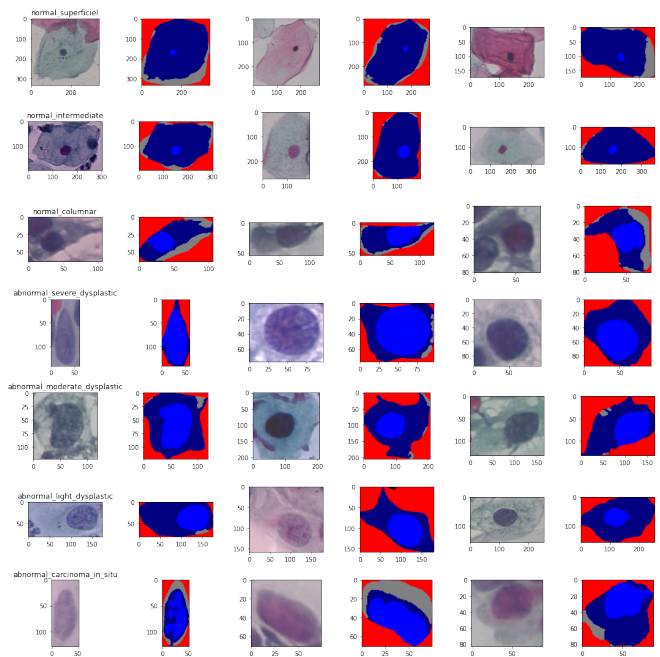
Cijeli podatkovni skup Herlev od 1834 slika nalazi se u folderu herlev_pap_smear te je preuzet sa stranice: <https://www.kaggle.com/umangjpatel/pap-smear-datasets>

SIPaKMeD

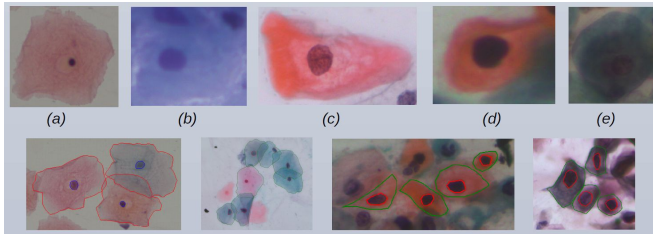
Ovaj skup podataka sadrži 5015 slika, od kojih je 966 višestaničnih slika dobivenih Papa-testom te je iz njih izrezano 4049 slika izoliranih stanica. Stanice su podijeljene na pet različitih klasa, ovisno o njihovoj staničnoj strukturi i morfološkim karakteristikama te je podjela vidljiva u Tablici 2.

Normalne stanice su skvamozne (pločaste) epitelne stanice čiji je tip definiran prema njihovom položaju u sloju epitela te stupnju sazrijevanja: superficialne-intermedijarne, parabazalne (Superficial-intermediate, Parabasal).

Abnormalne stanice ukazuju na postojanje pato-

**Slika 1:** Herlev: originalne i segmentirane normalne i abnormalne stanice.**Tablica 2:** Podatkovni skup SIPaKMeD

klasa	višestanične	izolirane
ABNORMALNE STANICE		
koilocitoza	238	825
diskeratoza	223	813
ukupan broj abnormalnih	461	1638
BENIGNE STANICE		
metaplastične stanice	271	793
ukupan broj benignih	271	793
NORMALNE STANICE		
superficialne-intermedijarne	126	831
parabazalne	108	787
ukupan broj normalnih	234	1618



Slika 2: *SIPaKMeD: pojedinačne stanice te višestanične slike uz označene granice*

loških svojstva te su karakterizirane morfološkim promjenama njihovih strukturnih dijelova. Infekcija HPV-om se očituje karakterističnim promjenama skvamoznih stanica od kojih su dvije patognomonične: koilocitoza i diskeratoza (Koilocytotic, Dyskeratotic).

Treća kategorija su benigne (Metaplastic) stanice koje predstavljaju transformacijsku zonu, odnosno područje u kojem se razvijaju gotovo svi predstadiji i stadiji raka vrata maternice. Njihova prisutnost u Papa-testu povezana je s većom stopom detekcije premalignih stanica te ih za potrebe klasifikacije možemo svrstati u klasu abnormalnih stanica.

Uz svaku višestaničnu sliku te uz svaku sliku pojedinačne izolirane stanice dobivene iz višestanične slike dane su datoteke s koordinatama piksela. Datoteke reprezentiraju granicu jezgre i granicu citoplazme svake od stanica.

Primjeri pojedinačnih stanica (prvi red) i višestaničnih slika sa vidljivim granicama jezgre i citoplazme (drugi red) dani su na Slici 2. Pojedinačne stanice su redom (a) superficijalna-intermedijarna, (b) parabazalna, (c) koilocitoza, (d) diskeratoza, (e) metaplastična stanica.

Cijeli podatkovni skup SIPaKMeD od 5015 slika te spomenutih datoteka s koordinatama granica se nalazi u folderima sipakmed_wsi_pap_smear (višestanične slike, engl. Whole Slide Images) i sipakmed_fci_pap_smear (pojedinačne izolirane stanice, engl. Full Cell Images) te je preuzet sa stranice: <https://www.kaggle.com/umangjpatel/pap-smear-datasets>

Cilj i hipoteze istraživanja problema

Cilj ovog istraživanja je napraviti precizan i efikasan model za detekciju premalignih stanica. To bi bio veliki napredak u odnosu na ručno provjeravanje koje zahtijeva stručno osoblje i više vremena zbog velikog broja slika.

Veličina jezgre stanice je dobar prediktor za premaligne stanice. Promatrajući Sliku 1, primjećujemo da je jezgra znatno manja u prva tri reda gdje su

stanice normalne, od jezgara u zadnja četiri reda gdje su stanice abnormalne.

Pregled dosadašnjih istraživanja

Neke metode koje su koristili drugi istraživači u rješavanju problema detekcije premalignih stanica je hijerarhijski clustering, Convolutional Neural Network, K-Nearest Neighbours, Artificial Neural Networks. Istraživanje koje je koristilo hijerarhijski algoritam je rezultiralo točnošću od 96%. Pri tome se pojavio problem sa izoliranjem jezgri stanica zbog šumova u teksturi. K-Nearest Neighbours i Artificial Neural Networks imaju manju točnost od hijerarhijskog algoritma te im točnost uvelike zavisi o vrsti stanica.

1. Genctav, S. Aksoy and Sevgen Onder: "Unsupervised Segmentation and Classification of Cervical Cell Images", 2012.
 - Neki korišteni feature-i: površina jezgre, svjetlina jezgre, najveći i najmanji promjer jezgre, opseg jezgre, omjer najduljeg i najkraćeg promjera, površina citoplazme...
 - Klasifikacija pomoću hijerarhijskog clustering-a koji rijetko pruža najbolja rješenja, ne radi dobro na velikim setovima podataka i pomiješanim tipovima podataka.
2. A. Kale and S. Aksoy: "Segmentation of Cervical Cell Images", International Conference on Pattern Recognition, 2010.
 - Algoritam: Hierarchical Segmentation algorithm
 - Jezgre nisu uspješno izdvojene iz nekih stanica zbog šuma u teksturi (noisy texture).
3. P. K. Malli, S. Nandyal: "Machine learning Technique for detection of Cervical Cancer using k-NN and Artificial Neural Network", International Journal of Emerging Trends Technology in Computer Science (IJETTCS), 2017.
 - K-Nearest Neighbours ima visoku cijenu izračunavanja (computation cost) i problem s određivanjem K.
 - Artificial Neural Network je imao slabu točnost (54%).
4. L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu et al., DeepPap: "Deep Convolutional Networks for Cervical Cell Classification", arXiv:1801.08616v1 [cs.CV], 2018.

- Ova metoda je imala dobar performans, ali klasifikacija uzima puno vremena. Par klasa stanica su krivo identificirani.
- Za metodu je potrebna informacija o centru jezgre.

Materijali, metodologija i plan istraživanja

Zbog neujednačenosti broja abnormalnih i normalnih stanica, po potrebi ćemo povećati naš skup podataka metodom rotacije, osne simetrije i drugim metodama. Obzirom da su slike iz Herlev skupa podataka različitih veličina, transformirat ćemo ih tako da budu iste veličine, koristeći funkcije u Python paketu scikit-image. Neki od feature-a koje ćemo uzeti u obzir su veličina jezgre, najveći i najmanji radijus, izduljenost jezgre (omjer najvećeg i najmanjeg radijusa).

Prva ideja je primijeniti logističku regresiju, promatrajući veličinu jezgri. Lako je moguće da će ovaj model biti prejednostavan i neće uspjeti.

Druga ideja je koristiti neuralne mreže s više ulaza kako bi stvorili moćniji i kompleksniji model koji će uzimati u obzir više feature-a.

Budući da je problem klasifikacijski, isprobat ćemo logističku regresiju jer je to najčešći klasifikacijski problem. Ukoliko nam točnost bude nezadovoljavajuća, isprobat ćemo nešto složeniji model pomoću konvolucijskih neuronskih mreža koje se često koriste u klasifikaciji slika.

Alati koje ćemo koristiti su Python i Tensorflow.

Očekivani rezultati predloženog projekta

Očekivani rezultat je model koji uspješno detektira premaligne stanice u nekom razumnom vremenu. Za ocjenjivanje rezultata ćemo koristiti već poznate ocjene klasificiranja: Accuracy, F1 score i Recall, s naglaskom na Recall zbog važnosti točnog detektiranja svih premalignih stanica. Naime, „False-Negative“ dijagnoza se smatra katastrofalnom greškom koju želimo izbjeći.

Literatura

- [1] Mithlesh A., Mittal N.: "Clustering Techniques on Pap-smear Images for the Detection of Cervical Cancer", 2017.
- [2] Yasha Singh, Dhruv Srivastava, Chandranand Anand: "Algorithms for screening of Cervical Cancer: A chronological review 1", 2018.
- [3] Genctav, S. Aksoy and Sevgen Onder: "Un-supervised Segmentation and Classification of Cervical Cell Images", 2012.
- [4] A. Kale and S. Aksoy: "Segmentation of Cervical Cell Images", International Conference on Pattern Recognition, 2010.
- [5] P. K. Malli, S. Nandyal: "Machine learning Technique for detection of Cervical Cancer using k-NN and Artificial Neural Network", International Journal of Emerging Trends Technology in Computer Science (IJETTCS), 2017.
- [6] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu et al., DeepPap: "Deep Convolutional Networks for Cervical Cell Classification", arXiv:1801.08616v1 [cs.CV], 2018.
- [7] Marina E. Plissiti, Christophoros Nikou, Giorgos Sfikas, Antonia Charchanti: "Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images", 2018.
- [8] Pliva zdravlje, <https://www.plivazdravlje.hr/aktualno/clanak/19771/Pap-test.html>
- [9] Hrvatski zavod za javno zdravstvo, <https://www.hzjz.hr/sluzba-epidemiologija-zarazne-bolesti/odjel-za-hiv-spolno-i-krvlju-prenosive-bolesti/vise-o-odjelu-za-hiv-spolno-i-krvlju-prenosive-zarazne-bolesti/>
- [10] World Health Organization, <https://www.who.int/health-topics/cervical-cancertab=tab1>