

Vilniaus Universitetas Matematikos ir informatikos institutas L I E T U V A



INFORMATIKOS INŽINERIJA (07 T)

MODELIS NAVIKŲ GENETINIŲ POKYČIŲ IR BIOŽYMENŲ RAIŠKOS DUOMENŲ INTEGRAVIMUI

Mindaugas Morkūnas

2017-10-21

Mokslinė ataskaita MII-DS-07T-17-

Matematikos ir informatikos institutas, Akademijos g. 4, Vilnius LT 08663 <u>http://www.mii.vu.lt/</u>

Turinys

1	Įva	ıdas	3
2	Ту	rimo motyvacija	3
3	Ek	sperimentai ir rezultatai	4
	3.1	Superpikselių klasifikavimas tradiciniais klasifikavimo algoritmais	4
	3.2	Vaizdo blokų klasifikavimas konvoliucinio neuroninio tinklo pagalba	6
4	lšv	ados	6
5	Lit	eratūra	7

1 Ivadas

Palyginti nesenas aukštos kokybės vaizdininimo technologiju pritaikymas šviesinėje mikroskopijoje pilno pjūvio patologijos vaizdams smarkiai išplėtė patologijos mokslo galimybes. Skaitmeniniuose vėžio patologijos vaizduose yra daug svarbių požymių tokių kaip vėžio ląstelių morfologija, vėžinio audinio mikroaplinka, kurie negali būti ištirti kitais įprastai naudojamais diagnostiniai metodais. Medicininių vėžio patologijos vaizdų analizė yra intensyvi, daug laiko ir aukšto lygio ekspertinių žinių reikalaujanti užduotis. Gydytojas patologas vertindamas patologiją turi atpažinti subtilius lastelių ir audinių pokyčius. Šią užduotį smarkiai apsunkina vėžinio audinio heterogeniškumas, atsirandantis dėl kloninės navikų evoliucijos, kuri yra būdinga daugumai vėžio tipų. Kiekvienas navikas yra genetiškai unikalus dėl skirtingo pavieniam naviko klone egzistuojančių mutacijų skaičius. Kitas heterogeniškumo lygmuo susijęs su baltymų sintezės reguliavimo mechanizmais. Tiek genų mutacijos, tiek biologiškai reikšmingų baltymų raiškos pokyčiai yra naudojami kaip prediktyviniai žymenys parenkant taikinių terapiją, todėl navikinio audinio heterogeniškumo problema yra kritinė individualizuotos vėžio terapijos plėtrai. Šiuolaikinėje patologijos analizėje plačiai taikomais imunohistocheminiais baltyminių žymenų raiškos metodais, galima atskleisti genų mutacijų efektą, bei šviesinės mikroskopijos pagalba tam tikromis biologinėmis savybėmis pasižyminčias naviko ląstelių subpopuliacijas galima atvaizduoti aukštos raiškos skaitmeniniuose vaizduose.

2 Tyrimo motyvacija

Skaitmeninėje patologijos analizėje viena iš pirmųjų užduočių ir dažnu atveju tarpinis tyrimo tikslas yra atskirų audinio grupių išskyrimas naviko vaizde (pvz.: epitelinis audinys / jungiamasis audinys). Tai padeda apibendrintai suvokti kur konkrečiai navike ir kokiu mastu konkretus biožymuo stebimas. Modernūs vėžiu sergančių pacientų stratifikavimo į prognostines ar predikcines grupes metodai vertina biožymens raišką kiekvienoje audinių grupėje (McLaughlin, J., 2016, Dunne, M.R., 2017).

Dauguma susijusių darbų remiasi įvairių požymių išgavimu iš skaitmeninio patologijos vaizdo (Gaboro filtrai, Haraliko tekstūra, Čebiševo momentai, spalvos/intensyvumo parametrai) ir naudoja gerai žinomus klasifikavimo algoritmus (Ahammer, H., 2009, Linder, N., 2012, Bianconi, F., 2015, Bunyak, F., 2015 ir Sethi, A., 2015, Nava, R., 2016). Požymiai paprastai išgaunami iš įvairiais būdais segmentuotų vaizdų. Naudojami stačiakampiai vaizdo blokai (Ahammer, H., 2009), persidengiantys stačiakampiai blokai (Linder, N., 2012), nevienodo dydžio stačiakampiai blokai (Bianconi, F., 2015), įvairūs superpikselių gavimo būdai (Bunyak, F., 2015, Sethi, A., 2015, Bejnordi, B. E., 2016). Istoriškai tokių metodų efektyvumas buvo didinamas išskiriant daugiau ir įvairesnių vaizdo požymių, tačiau keletą pastarųjų metų dažniausiai naudojami gilaus apmokymo konvoliuciniai neuroniai tinklai nereikalaujantys jokio požymių išskyrimo (Huang, Y., 2015, Sethi, A., 2015, Litjens, G., 2016, Xu, J., 2016, Huang, Y., 2017).

Tiek tradiciniams tiek gilaus apmokymo klasifikavimo algoritmams reikalingi tiksliai anotuoti apmokymo duomenys. Dauguma susijusių darbų yra ribojami pilno pjūvio vaizdų anotavimo etape, kadangi tai yra ilgas procesas reikalaujantis aukšto ekspertinių žinių lygio. Dėl skubos, ar dėl to jog anotavimui naudojami ne pačios aukščiausios raiškos vaizdai gaunamos netikslios ar neteisingos anotacijos. MII-DS-07T-17-3

Įprastai audinių grupių išskyrimas navikų vaizduose turi būti adaptuojamas atsižvelgiant į labai specifinius galutinius analizės tikslus - ar tai būtų kito tipo vėžinio susirgimo tyrimas, ar naujų anksčiau netirtų naviko savybių tyrimas.

3 Eksperimentai ir rezultatai

Šioje ataskaitoje aprašomame metode superpikseliai naudojami ne tik požymių išskyrimui, bet ir etaloninių duomenų gavimui tiek tradiciniams klasifikavimo algoritmams, tiek ir konvoliucinio tinklo mokymui. Metodas pritaikytas trims storosios žarnos vėžio vaizdams.

3.1 Superpikselių klasifikavimas tradiciniais klasifikavimo algoritmais

Superpikselių klasifikavimo algoritmo seka pateikiama sekančioje schemoje:

1. Gydytojo patologo parinktos naviko zonos pilno pjūvio vaizduose sudalinamos į superpikselius naudojant SLIC algoritmą (Achanta, R., 2011).



1 pav. 1000 x 2000 pikselių naviko zona yra dalinama į apytikriai 350 superpikselių.

2. Superpikselių kontūrai atvaizduojami naviko vaizde ir anotuojami eksperto (kiekvienas superpikselis priskiriamas vienai iš naviko audinio grupių, fonas pašalinamas kaip neinformatyvi vaizdo vieta).

	Epitelinis audinys	Jungiamasis audinys
Apmokymo rinkinys	28000	28000
Validavimo rinkinys	2000	2000
Tikrinimo rinkinys	7068	3929

1	lentelė	super	niksel	in 1	rinl	kin	ia	i
1	fentere.	Super	pinsei	цц	um	vin	Iu	1.

3. Superpikseliai požymių išskyrimui paruošiami trim būdais - naudojajmi dvimačiai spalvoti (RGB) vaizdo blokai aprėminantys kiekvieną superpikselį, analogiški

dvimačiai pilkos spalvos atspalvių vaizdo blokai ir "ištiesinti" superpikseliai, kur viena dimensija gaunama visus konkretaus superpikselio pikselius saugant viename vektoriuje, o antra dimensija gaunama iš kiekvieno spalvos kanalo.

4. Iš kiekvieno superpikselio išskiriami 13 tekstūros požymių (Haralick, R.M., 1979) ir 8 spalvos požymiai (vidutinė pikselio vertė ir dispersija kiekvienam spalvos kanale atskirai ir visiems kartu).

5. Naudojant pagrindinių komponenčių analizę sumažinamas superpikselių požymių rinkinių dimensijų skaičius.

6. Superpikseliai klasifikuojami pagal išskirtus požymius naudojant skirtingus klasifikavimo algoritmus:

- RDF atsitiktinių miškų klasifikatorius (su 25 bazinio lygmens medžiais);
- SVM atraminių vektorių mašina (naudotas C-SVC algoritmas su radialine bazine funkcija);
- MLP daugiasluoksnis perceptronas (su dviem paslėptais neuronų sluoksniais atitinkamai sudarytais iš 1000 ir 40 neuronų);
- DL gilaus mokymo tinklas (su trim paslėptais sluoksniais iš atitinkamai 1024, 512 ir 128 neuronų).

Gauti klasifikavimo rezultatai palyginti su etaloniais duomenimis. Klasifikavimo tikslumas vertintas apskaičiuojant plotą po ROC kreive (AUC).

			Požymiai	
Superpikslelių paruošimo būdas	Klasifikavimo algoritmas	Spalva	Spalva ir tekstūra	Tekstūra
	RDF	0,9569	0,9577	0,9367
2D "ištiesinti"	SVM	0,9611	0,9613	0,9498
superpikseliai	MLP	0,9668	0,9660	0,9591
	DL	0,9608	0,9580	0,9397
	RDF	0,9466	0,9533	0,9385
2D RGB	SVM	0,9515	0,9620	0,9488
superpikseliai	MLP	0,9577	0,9666	0,9542
	DL	0,9556	0,9575	0,9449
	RDF	0,7083	0,8553	0,8694
2D pilki	SVM	0,7817	0,8688	0,8861
superpikseliai	MLP	0,7967	0,8765	0,8885
	DL	0,7967	0,8694	0,8777

2 lentelė. Superpikselių klasifikavimo tikslumas išreikštas kaip AUC.

Rezultatai pateikti 2 lentelėje rodo, kad naudojant sąlyginai nedaug spalvos ir tekstūros požymių galima pasiekti aukštą klasifikavimo tikslumą. Tikslumas ženkliai sumažėja atsisakius spalvos požymių (tiek naudojant pilkos spalvos atspalvių superpikselius, tiek klasifikavimui naudojant vien tik tekstūros požymius).

3.2 Vaizdo blokų klasifikavimas konvoliucinio neuroninio tinklo pagalba

Klasifikavimo algoritmo naudojant konvoliucinį neuroninį tinklą schema:

1-2 žingsniai tokie patys kaip ir superpikslelių klasifikavimo atveju (žr. 3.1).

3. Anotuotų superpikselių viduje parenkami stačiakampiai vaizdo blokai perstumiant 72x72 pikselių dydžio rėmelį aplink superpikselio masės centrą. Iš kiekvieno superpikselio gaunami 5 persidengiantys vaizdo blokai.



2. pav. vaizdo blokų parinkimas konvoliucinio tinklo apmokymui

	Epitelinis audinys	Jungiamasis audinys
Apmokymo rinkinys	40000	40000
Validavimo rinkinys	25000	5000
Tikrinimo rinkinys	26000	6000

3 lentelė. vaizdo blokų rinkiniai.

4. Gauti vaizdo blokai klasifikuojami naudojant skirtingų architektūrų konvoliucinį neuroninį tinklą (4 lentelė).

conv2d_1	conv2d_2	conv2d_3	conv2d_4	dense_1	AUC		
32	32	64	64	384	0,9704		
32	32	64	64	1024	0,9667		
64	64	128	128	384	0,9727		
64	64	128	128	1024	0,9687		
96	96	192	192	384	0,9745		
96	96	192	192	1024	0,9721		

4 lentelė. Skirtingų architektūrų konvoliucinio neuroninio tinklo klasifikavimo tikslumas išreikštas kaip AUC.

Iš 4 lentelėje pateiktų rezultatų matyti, kad konvoliucinio neuroninio tinklo klasifikavimo tikslumas didėja didinant neuronų skaičių konvoliuciniuose sluoksniuose ir yra aukštesnis nei pasiekiamas klasifikuojant superpikselius.

4 Išvados

Aukščiausias superpikselių klasifikavimo tikslumas pasiektas naudojant daugiasluoksnį perceptroną spalvotiems superpikseliams (AUC = \sim 0.967). Kitų autorių pasiekti rezultatai tam pačiam vėžio tipui yra labai artimi (AUC = 0,97 - 0,99 - Linder, N., 2012 ir Nava, R., 2016).

Aukščiausias pasiektas konvoliucinio neuroninio tinklo tikslumas (AUC = \sim 0.975, 2 lentelė) yra artimas kitų autorių publikuotiems rezultatams gautiems naudojant kitų

MII-DS-07T-17-

tipų navikų duomenis - krūties vėžio (AUC = 0.9316 - Xu, J., 2016) ir prostatos vėžio (AUC = 0.965 - Sethi, A., 2015 ir AUC = 0.99 - Litjens, G., 2016).

5 Literatūra

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012) SLIC Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274--2282;

Ahammer, H., Kriipfl, J., Hackl, C., Sedivy, R., (2009). Image statistics and data mining of anal intraepithelial neoplasia. Pattern Recognition Letters, 29:2189–2196;

Bejnordi, B. E., Balkenhol, M., Litjens, G., Holland, R., Bult, P., Karssemeijer, N., van der Laak, J.A. (2016). Automated Detection of DCIS in Whole-Slide H&E Stained Breast Histopathology Images. IEEE TRANSACTIONS ON MEDICAL IMAGING, 35(9), 2141--2150;

Bianconi, F., Alvarez-Larrn, A., Fernndez, A. (2015). Discrimination between tumour epithelium and stroma via perception-based features. Neurocomputing, 154(0):119–126;

Bunyak, F., Hafiane, A., Al-Milaji, Z., Ersoy, I., Haridas, A., Palaniappan, K. (2015). A segmentation-based multi-scale framework for the classification of epithelial and stromal tissues in H&E images. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), DOI: 10.1109/BIBM.2015.7359726;

Dunne, M.R., Michielsen, A.J., O'Sullivan, K.E., Cathcart, M.C., Feighery, R., Doyle, B., Watson, J.A., O'Farrell, N.J., Ravi, N., Kay, E., Reynolds, J.V., Ryan, E.J., O'Sullivan, J. (2017). HLA-DR expression in tumor epithelium is an independent prognostic indicator in esophageal adenocarcinoma patients. Cancer Immunology, Immunotherapy, 66(7), 841–850. http://doi.org/10.1007/s00262-017-1983-1;

Haralick, R.M. (1979). Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5), 786--804;

Huang, Y., Zheng, H., Liu, C., Ding, X., Rohde, G. (2015). Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. IEEE Journal of Biomedical and Health Informatics, DOI: 10.1109/JBHI.2017.2691738;

Linder, N., Konsti, J., Turkki, R., Rahtu, E., Lundin, M., Nordling, S., Haglund, C., Ahonen, T., Pietikäinen, M., Lundin, J. (2012). Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. Diagnostic Pathology, 7, 22, http://doi.org/10.1186/1746-1596-7-22;

Litjens, G., Sanchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen - van de Kaa, C., Bult, P., van Ginneken, B., van der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific Reports, 6:26286, DOI: 10.1038/srep26286;

McLaughlin, J., Han, G., Schalper, K.A., Carvajal-Hausdorf, D., Pelakanou, V., Rehman, J., Velcheti, V., Herbst, R., LoRusso, P., and Rimm, D.L. (2016). Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-small Cell Lung Cancer (NSCLC). JAMA Oncology, 2(1), 46–54. http://doi.org/10.1001/jamaoncol.2015.3638;

MII-DS-07T-17-

Nava, R., Gonzalez, G., Kybic, J., Escalante-Ramirez, B. (2016). Classification of Tumor Epithelium and Stroma in Colorectal Cancer Based on Discrete Tchebichef Moments. In: Oyarzun Laura C. et al. Clinical Image-Based Procedures. Translational Research in Medical Imaging. CLIP 2015. Lecture Notes in Computer Science, vol 9401. Springer, Cham;

Sethi, A., Sha, L., Vahadane, A.R., Deaton, R.J., Kumar, N., Macias, V., Gann, P.H. (2016). Empirical comparison of color normalization methods for epithelial-stromal classification in H and E images. Journal of Pathology Informatics, 7, 17. http://doi.org/10.4103/2153-3539.179984;

Xu, J., Luo, X., Wang, G., Gilmore, H., Madabhushi, A. (2016). A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing, 191, 214–223. http://doi.org/10.1016/j.neucom.2016.01.034.