



**Vilniaus universitetas
Duomenų mokslo ir skaitmeninių
technologijų institutas
L I E T U V A**



INFORMATIKOS INŽINERIJA (07 T)

**MODELIS NAVIKŲ GENETINIŲ
POKYČIŲ IR BIOŽYMENŲ RAIŠKOS
DUOMENŲ INTEGRAVIMUI**

Mindaugas Morkūnas

2018 m. spalio

Mokslinė ataskaita DMSTI-DS-07T-18-2

VU Duomenų mokslo ir skaitmeninių technologijų institutas, Akademijos g. 4,

Vilnius LT-08663

www.mii.lt

Santrauka

Kolageno gijos yra paprasčiausi jungiamojo audinio elementai. Subtili šio audinio struktūra pakinta esant įvairioms žmogaus organizmo patologijoms, o ypač - tvirtų audinių navikuose. Todėl, kolageno gijų atpažinimas ir detalus kiekybinis charakterizavimas yra itin svarbus įvairių patologinių būklių klasifikavimui. Kolageno struktūros biologinių audinių preparatuose vizualizavimui naudojami labai įvairūs metodai. Daugelis jų (tokie kaip - elektroninė mikroskopija, atominės jėgos mikroskopija, antros harmonikos generacijos mikroskopija) yra apriboti tik specializuotam naudojimui dėl dažnai aukštos įrangos kainos, sudėtingo mėginio paruošimo, ir dažnai - mažo matymo lauko. Šiame kontekste, gerokai pigesnė ir labiau įprasta šviesinė mikroskopija yra ignoruojama, nes negali pasiūlyti tokio detalumo lygio kaip anksčiau paminėtos specializuotos mikroskopijos rūšys.

Šiame tyrime, apmokėme pilnai konvoliucinį gilaus mokymo neuroninį tinklą atpažinti krūties vėžio navikų kolageno struktūrą šviesinės mikroskopijos vaizduose. Kolageno struktūrai charakterizuoti išgauti parametrai buvo validuoti kliniškai nedidelėje krūties vėžio pacientų imtyje. Sukurtas modelis be jokios specifinės adaptacijos buvo pritaikytas kolageno struktūrai atpažinti ir kitų tipų navikų vaizdams, bei nevėžinių patologinių būklių vaizdams.

Šis modelis leis tiksliau charakterizuoti tiriamų navikų ląstelių populiacijas įvertinant jų mikroaplinką, sąveiką su kitais navikinio audinio komponentais, organizmo imunine sistema.

Reikšminiai žodžiai: navikų mikroaplinka, kolageno gijos, šviesaus lauko mikroskopija, gilaus mokymo neuroninis tinklas.

Turiny

1	Įvadas	4
2	Tyrimo motyvacija	5
3	Eksperimentai ir rezultatai	5
3.1	„U-Net“ tinklo architektūra	5
3.2	Etaloniniai duomenys	6
3.3	Modelio schema	6
3.4	Kolageno struktūros požymiai	7
3.5	Klinikinis kolageno struktūros požymių validavimas	8
3.6	Taikymo išplėtimas	10
3.7	Išvados.....	11
4	Literatūra.....	12
	Priedai	13

1 Įvadas

Skaitmeniniai vėžio patologijos vaizdai turi daug vėžinio susirgimo diagnozei svarbių požymių, tokių kaip vėžio ląstelių morfologija, ar vėžinio audinio mikroaplinka, kurių subtilius pokyčius turi atpažinti gydytojas patologas. Šią užduotį smarkiai apsunkina vėžinio audinio heterogeniškumas, atsirandantis dėl kloninės navikų evoliucijos, kuri yra būdinga daugumai vėžio tipų [1]. Kiekvienas navikas yra genetiškai unikalus dėl skirtingo pavieniam naviko klone egzistuojančių mutacijų skaičius. Dėl šios priežasties net ir to paties naviko ląstelės gali turėti skirtingą morfologiją ir fenotipą, bei skirtingai sąveikauti su naviko mikroaplinka. Yra žinoma, jog besiformuodamas navikas aktyviai moduliuoja sau palankią mikroaplinką: suformuoja savo kraujagysles, perorganizuoja kolageno gijų tinklą, sutrikdo imuninę sistemą, užsitikrina kelius plitimui. Naujausi tyrimai rodo šios ypatingos sąveikos svarbą bendram naviko piktybiniam potencialui ir itin didelę įtaką daugelio priešvėžinių terapijų efektyvumui [2, 3].

Navikų mikroaplinka (angl. *tumor microenvironment* (TME)) tai aplinka, kurioje navikas egzistuoja. Į ją įeina aplinkinės kraujagyslės, įvairios imuninės ir uždegiminės ląstelės, jungiamojo audinio ląstelės (fibroblastai), signalinės molekulės ir tarpląstelinė medžiaga. Pastarasis naviko mikroaplinkos komponentas (angl. *extracellular matrix* (ECM)) didžiąja dalimi sudarytas iš baltymo kolageno gijų ir sudaro specifinę erdvinę struktūrą - tinklą, kuriame įsitvirtina naviko ląstelės. Ši subtili struktūra pakinta esant įvairioms žmogaus organizmo patologijoms, o ypač - tvirtųjų audinių navikuose. Todėl, kolageno gijų atpažinimas ir detalus kiekybinis charakterizavimas yra itin svarbus įvairių patologinių būklių klasifikavimui.

Kolageno struktūros vizualizavimui navikų (ir bendrai visų žmogaus audinių) preparatuose naudojami labai įvairūs metodai - antros harmonikos generacijos [4], poliarizuotos šviesos ir skystų kristalų poliarizuotos šviesos [5], elektroninė ir skenuojanti elektroninė [6], atominės jėgos mikroskopijos [7]. Dažnu atveju tai brangi ir sudėtinga įranga reikalaujanti ypatingo audinių paruošimo, ar net patyrusio optikos inžinieriaus darbo.

2 Tyrimo motyvacija

Skaitmeninis patologijos vaizdų apdorojimas ir analizė patologijoje nėra naujiena. Pirmieji darbai skaitmeninės patologijos srityje siekia 90-uosius metus [8,9]. Šiame tyrimų lauke paskutinius keletą metų itin efektyviai naudojami neuroniniai tinklai [10]. Palyginti pigi ir medicinoje įprasta šviesinė mikroskopija taip pat naudojama kolageno struktūrai charakterizuoti [11], tačiau dėl nepakankamo detalumo lygio kiekybiniai kolageno gijų matavimai šviesinėje mikroskopijoje apribotai rankiniu tam tikrų požymių anotavimu ir skaičiavimu, bei labai apibendrintu, kategoriniu rezultatų vertinimu. Rankinis ir dėl to lėtas aukštos kvalifikacijos mediko darbas šį kolageno struktūros vertinimo būdą padaro taip pat ir brangiu. Tenka konstatuoti, jog šiandien nėra publikuotų metodų leidžiančių automatizuotai išgauti kolageno struktūrą iš šviesinės mikroskopijos navikų patologijos vaizdų.

3 Eksperimentai ir rezultatai

Šiame tyrime, apmokėme pilnai konvoliucinį gilaus mokymo neuroninį tinklą atpažinti navikų kolageno struktūrą šviesinės mikroskopijos vaizduose. Kolageno struktūrai charakterizuoti išgauti parametrai buvo validuoti kliniškai. Sukurtas modelis be jokios specifinės adaptacijos buvo pritaikytas kolageno struktūrai atpažinti ir kitų tipų navikų vaizdams, bei nevėžinių patologinių būklių vaizdams.

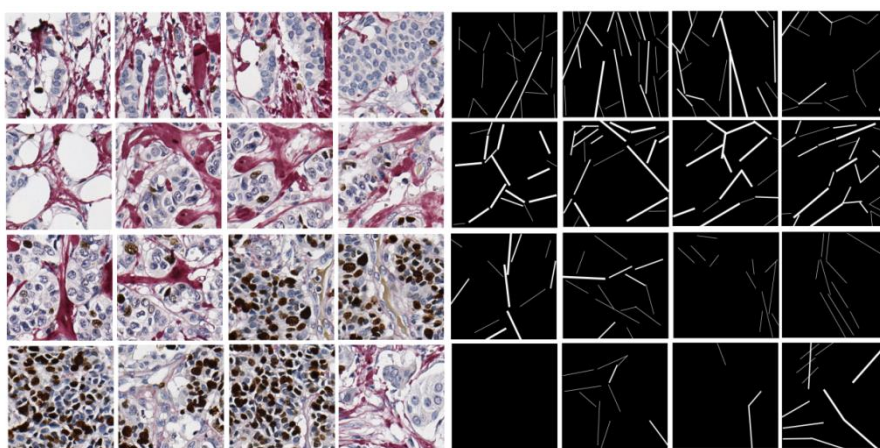
3.1 „U-Net“ tinklo architektūra

„U-Net“ yra konvoliucinis neuroninis tinklas sukurtas biomedicininėms vaizdų segmentavimui. Tinklo architektūra yra pilnai konvoliucinė - jame nėra nei vieno tankiai sujungto sluoksnio. Pagrindinė tinklo ypatybė ta, jog jis susideda iš dviejų dalių - užkoduojančios ir iškoduojančios. Ši U raidės pavidalo tinklo forma lėmė ir pavadinimą. Užkoduojanti tinklo dalis yra įprastas konvoliucinis tinklas, kuriame vaizdo erdvinė informacija „suspaudžiama“ dėl pakartotinai taikomų konvoliucijų, o požymių erdvė išplečiama. Iškoduojanti tinklo dalis tai seka transponuotų konvoliucijų ir apjungimų su lygiagrečių užkoduojančios tinklo dalies sluoksnių

požymiais. Tokia tinklo architektūra leidžia dirbti su mažomis mokymo vaizdų aibėmis [12].

3.2 Etaloniniai duomenys

Tinklui apmokinti naudoti 256 x 256 pikselių dydžio anotuoti krūties vėžio navikų vaizdo blokai. Kolageno struktūrai išryškinti pritaikytas specifinis histologinio preparato dažymo metodas „Sirius-Red“ dažas. Šis dažas kolageno gijas nudažo ryškiai raudonai (1.pav).

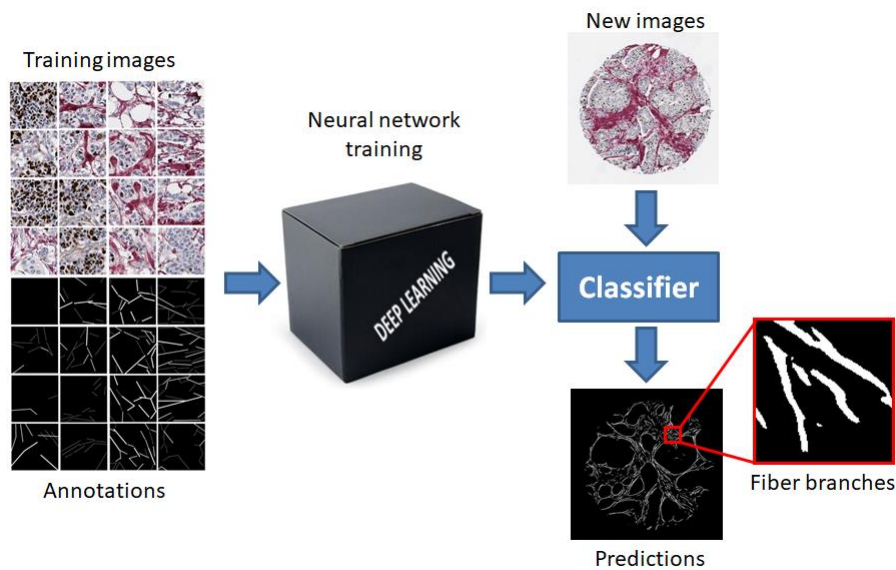


1 pav. Etaloniniai vaizdai ir kolageno anotacijos.

Viso anotuota 116 vaizdo blokų. Apmokymo vaizdų imtis išplėsta iki 464 blokų taikant vaizdo pasukimą.

3.3 Modelio schema

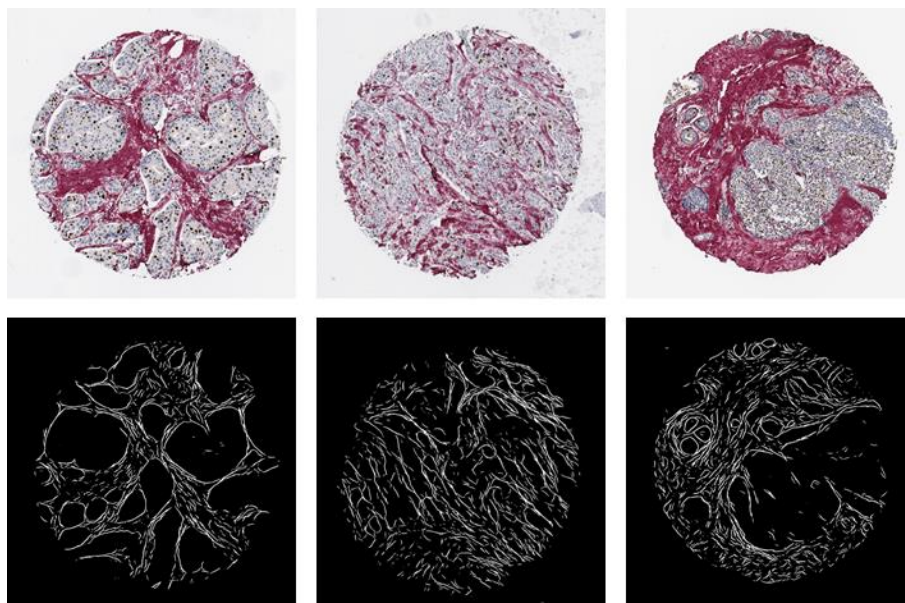
Bandymų metu buvo parinkti tinkamiausi modelio parametrai - „elu“ (eksponentinė linijinė) aktyvacijos funkcija tarp visų konvoliucijų, ir sigmoidinė funkcija rezultatų išvedimo sluoksniui. Apsimokymo efektyvumas buvo stebimas pagal „Jackard“ indeksą, mokymo paklaida matuojama binarine krosentropija. Neuroninis tinklas buvo mokomas 1000 epochų, tačiau maksimalų tikslumą pasiekė jau po 300 ir buvo užfiksuotas šioje būsenoje. 2 paveiksle pateikiama principinė viso modelio schema.



2 pav. Principinė modelio schema.

3.4 Kolageno struktūros požymiai

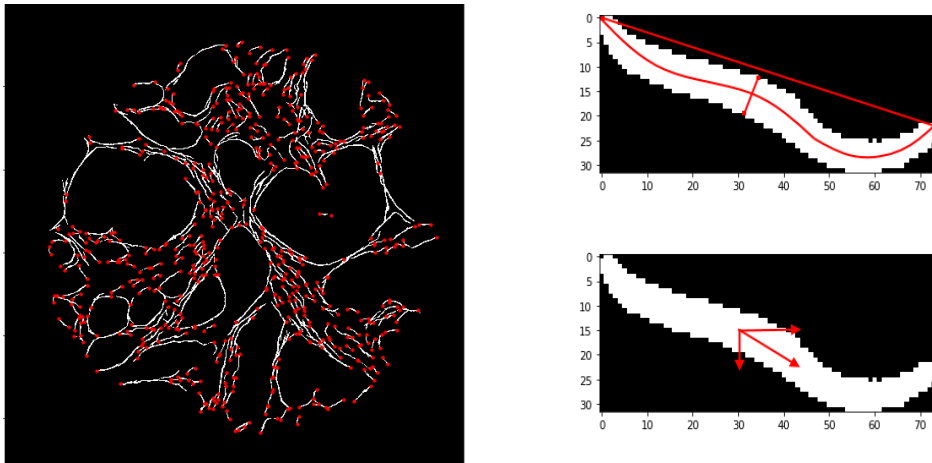
Po neuroninio tinklo apmokymo, buvo išanalizuoti 93 krūties vėžio navikų vaizdai (3 pav, didesnės raiškos vaizdas - 1S pav. priede).



3 pav. Originalūs (viršuje) ir neuroniniu tinklu išanalizuoti (apačioje) krūties vėžio navikų vaizdai.

Kolageno struktūros vaizdams buvo skaičiuojami įvairūs parametrai grubiai suskirstyti į grupes - gijų geometrijos parametrai, gijų krypties parametrai, fraktalinės

charakteristikos, tekstūros charakteristikos. Viso kiekvienam analizuotam vaizdui buvo išgaunami 24 parametrai (4 pav. ir 1S lentelė priede).

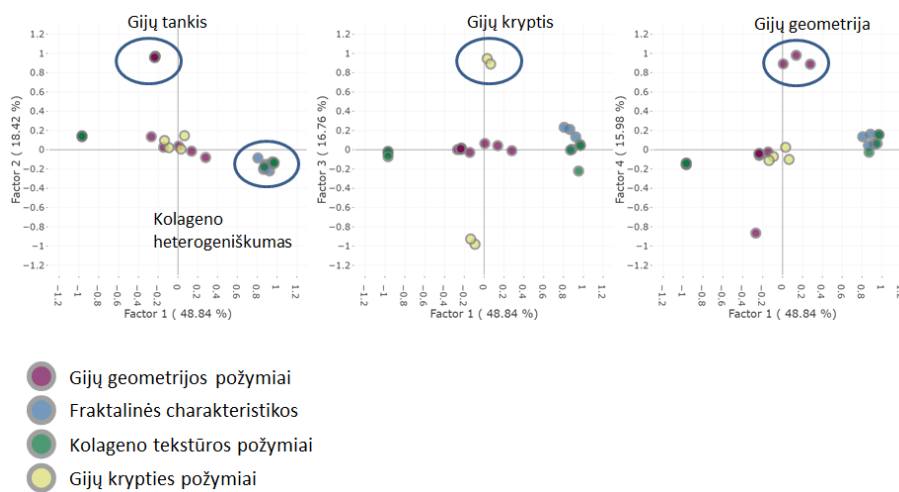


4. pav. Kolageno struktūros požymiai.

Šiame tyrime kolageno gijomis vadiname neuroninio tinklo sugrąžintus atskirus objektus - tai nėra tikrų biocheminių kolageno gijų ekvivalentas, greičiau tai vienodai orientuotų erdvėje arti esančių tikrųjų kolageno gijų junginiai (literatūroje vadinami kolageno šakomis - angl. *fiber branches*), tačiau patogumo dėlei toliau naudojame „gijų“ sąvoką.

3.5 Klinikinis kolageno struktūros požymių validavimas

Faktorinė kolageno struktūros vaizdų požymių analizė atskleidė keturis pagrindinius faktorius paaiškinančius variaciją duomenų rinkinyje (5 pav.):



5. pav. Faktorinė kolageno struktūros vaizdų požymių analizė.

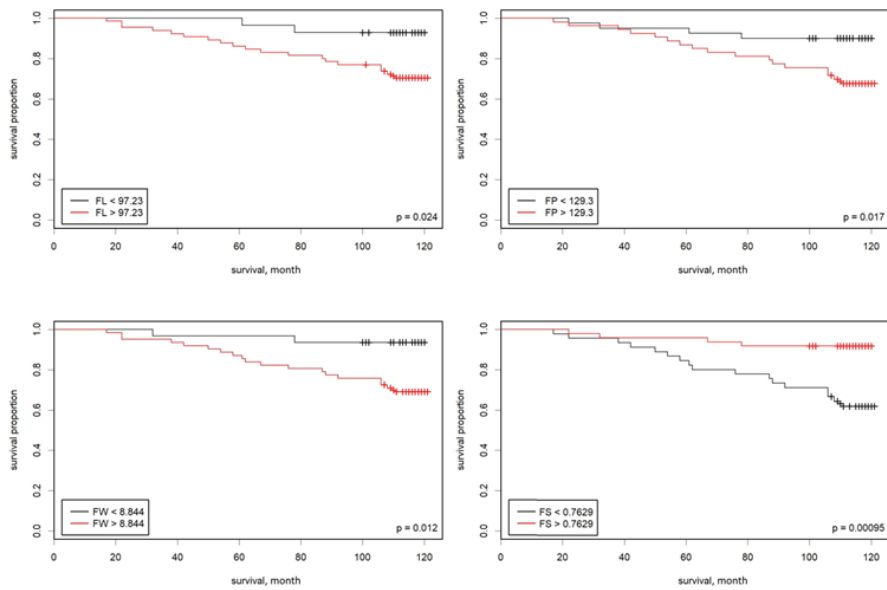
- su pirmamu faktoriumi stipria teigiama koreliacija pasižymėjo tekstūros požymiai ir fraktalinės charakteristikos. Šių požymių visuma atspindi kolageno struktūros heterogeniškumą;
- su antru faktoriumi stipria teigiama koreliacija pasižymėjo visi apskaičiuoti kolageno gijų tankio požymiai (n = 2,4,8,16);
- su trečiu faktoriumi stipria teigiama koreliacija pasižymėjo gijų krypties variabilumo požymiai, o stipria neigiama koreliacija pasižymėjo gijų krypties vidutinės reikšmės. Šių požymių visuma geriausiai apibūdina atvejus, kuriuose gausu trumpų įvairios krypties kolageno gijų;
- ketvirtą faktorių geriausiai nusako kolageno gijų geometrijos požymiai. Šių požymių visuma atspindi procesą vykstantį neuroniniam tinklui jungiant kolageno gijas į „šakas“ - kuo daugiau tikrųjų kolageno gijų tinklas mato arti erdvėje, tuo į platesnes ir ilgesnes šakas gali jas apjungti. Stipria neigiama koreliacija su ketvirtu faktoriumi pasižymėjo kolageno gijų tiesumo parametras.

Faktorizuoti kolageno struktūros vaizdų požymiai buvo tirti Kokso proporcingo pavojingumo modelyje (angl. *Cox proportional hazards model*):

1 lentelė. Kokso proporcingo pavojingumo modelis.

	Pavojingumo santykis	p-vertė	95% patikimumo intervalas	
			apatinė riba	viršutinė riba
Paciento amžius	1.047	0.0240	1.006	1.090
“Gijų geometrija” (4 faktorius)	1.200	0.0148	1.036	1.389
“Kolageno heterogeniškumas” (1 faktorius)	0.951	0.0495	0.905	0.999

Kokso modelis atskleidžia nepriklausomus veiksniai patikimai nusakančius ligos eigą. Matome, jog aukštesnės kolageno gijų geometrinių požymių vertės - platesnės, ilgesnės ir stipriau išlinkusios neuroninio tinklo atpažintos kolageno gijos („šakos“) sietinos su blogesne ligos baigtimi.

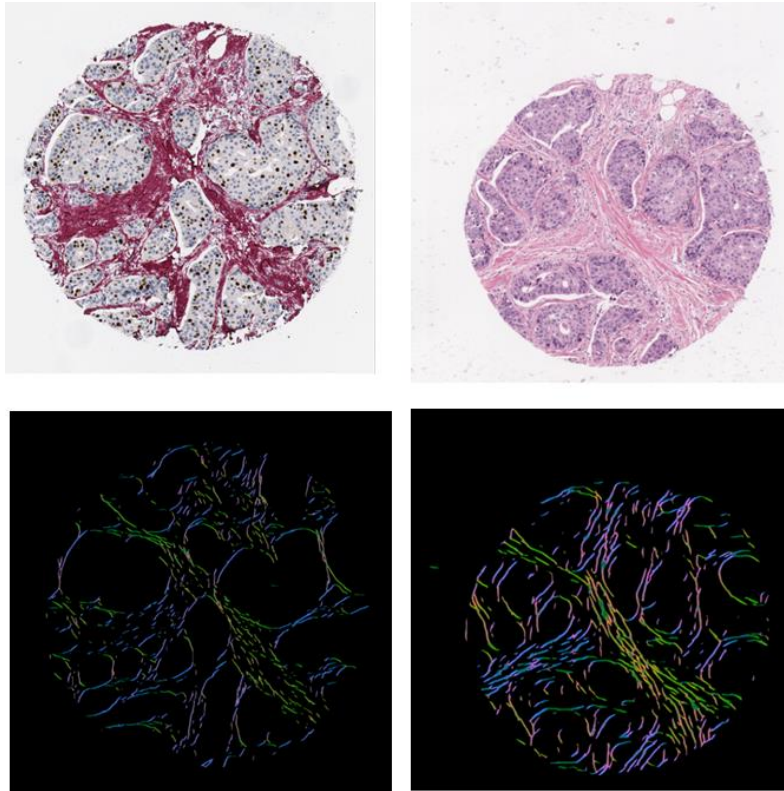


6 pav. Krūties vėžio pacientų stratifikavimas į geros ir blogos prognozės grupes pagal kolageno gijų geometrijos požymius.

Visiems kolageno gijų geometriją nusakantiems požymiams (imant kiekvieną jų atskirai) galėjome surasti slenkstines vertes, leidžiančias pacientus patikimai stratifikuoti į geros arba blogos prognozės grupes (6 pav.)

3.6 Taikymo išplėtimas

Neuroninis tinklas apmokintas atpažinti kolageno struktūrą krūties vėžio navikų vaizduose, be jokios papildomos adaptacijos, be papildomo apmokymo buvo išbandytas tiek su skirtingų patologijų, tiek ir skirtingais histocheminiais dažais nudažytų preparatų vaizdais - kito vėžio tipo navikų vaizdais (limfomos ir storosios žarnos vėžio), sveikos ragenos vaizdais, inkstų transplantų audinio vaizdais (7 pav. ir 2S, 3S, 4S, 5S paveikslai priede).



7 pav. Du to paties krūties vėžio naviko mikropreparatai nudažyti skirtingais dažais - „Sirius-Red“ ir „HE“ (viršuje, iš kairės į dešinę) ir atitinkamai jų kolageno struktūros atpažintos to paties U-net modelio (apačioje).

3.7 Išvados

Gilaus mokymo neuroninis tinklas gali būti efektyviai išmokintas atpažinti kliniškai reikšmingą kolageno struktūrą šviesinės mikroskopijos vaizduose. Tai pilnai automatinis ir pigia vaizdinimo technologija paremtas kolageno struktūros analizės metodas.

Metodo lankstumas leidžia jį naudoti ne tik navikinių patologijų tyrimuose ir diagnostikoje, bet ir pritaikyti kitų su kolageno pokyčiais susijusių patologijų tyrimuose.

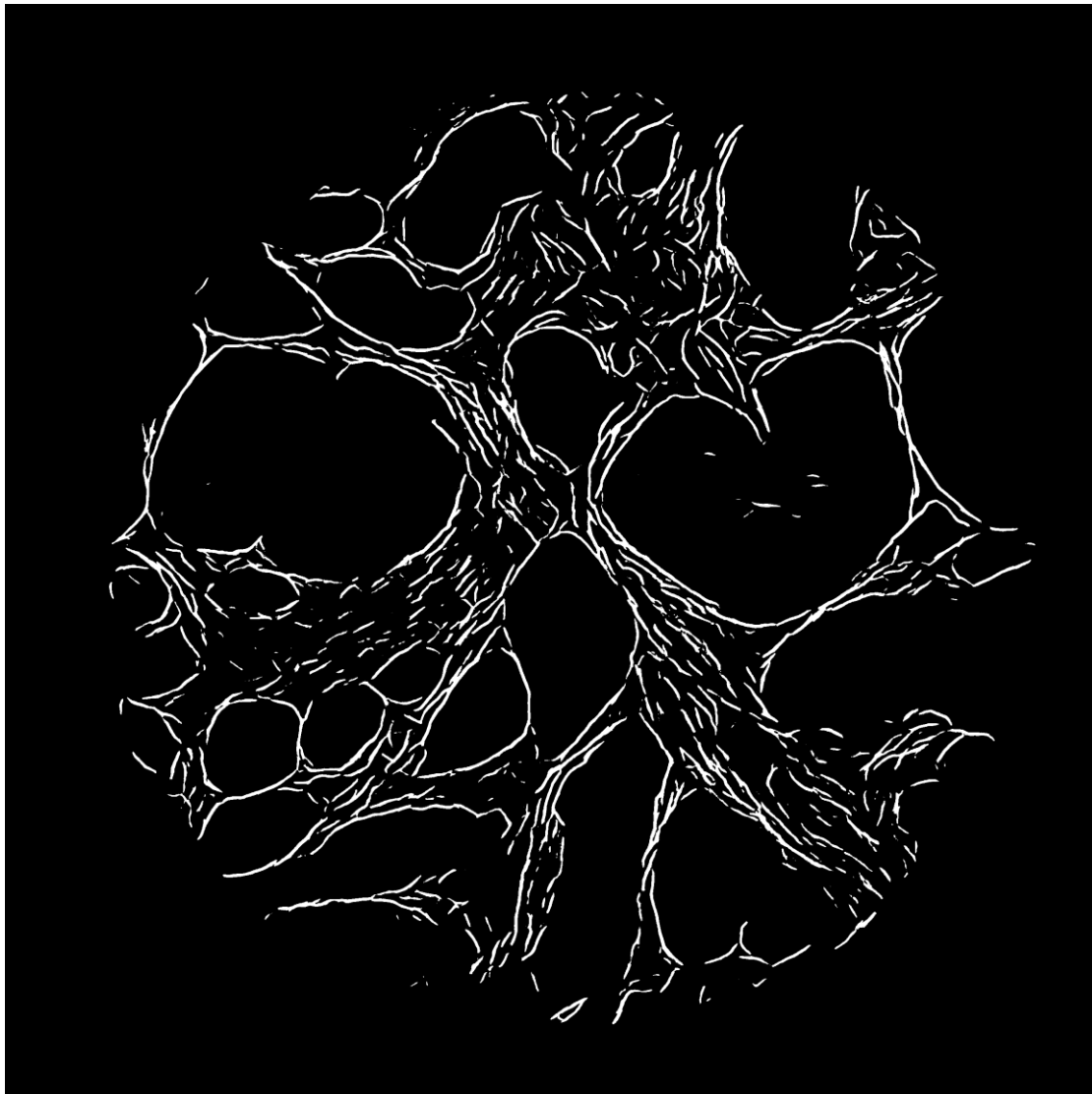
4 Literatūra

1. Greaves, M. and C. C. Maley (2012). “Clonal evolution in cancer”. *Nature* 481: 306.
2. Korneev, KV et al. (2017). “TLR-signaling and proinflammatory cytokines as drivers of tumorigenesis”. *Cytokine*. 89: 127–135.
3. Joyce, J et al. (2015). “T cell exclusion, immune privilege, and the tumor microenvironment”. *Science Magazine*. pp. 74–80.
4. Hompland, T et al. (2008). “Second-harmonic generation in collagen as a potential cancer diagnostic parameter”. *J Biomed Opt.* 13(5):054050
5. Manjunatha, BS et al. (2015). “Histopathological evaluation of collagen fibers using picosirius red stain and polarizing microscopy in oral squamous cell carcinoma”. *J Cancer Res Ther.* 11(2):272-6.
6. Kobayashi, M (2006). “Scanning electron microscopic study of the three-dimensional structure of the collagen sheath surrounding cancer cells after single high-dose irradiation”. *Med Mol Morphol.* 39(2):106-12.
7. Doss, BL et al. (2015). “AFM Indentation Reveals Actomyosin-Based Stiffening of Metastatic Cancer Cells during Invasion into Collagen I Matrices” *Biophysical journal.* 108(2):142A.
8. Bartels, PH et al. (1992). “Bayesian belief networks in quantitative histopathology,” *Anal. Quant. Cytol. Histol.*, vol. 14, no. 6, pp. 459–473, 199
9. Thiran, JP and Macq, B. (1996). “Morphological feature extraction for the classification of digital images of cancerous tissues,” *IEEE Trans. Biomed. Eng.*, vol. 43, no. 10, pp. 1011–1020.
10. Madabhushi, A and Lee, G. (2016). “Image analysis and machine learning in digital pathology: Challenges and opportunities”. *Med Image Anal.* 33: 170–175.
11. Dekker, TJA et al. (2015) “Disorganised stroma determined on pre-treatment breast cancer biopsies is associated with poor response to neoadjuvant chemotherapy: Results from the NEOZOTAC trial” *Mol Oncol.* 9(6): 1120-1128.
12. Ronneberger, O et al. (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, LNCS, Vol.9351: 234--241.

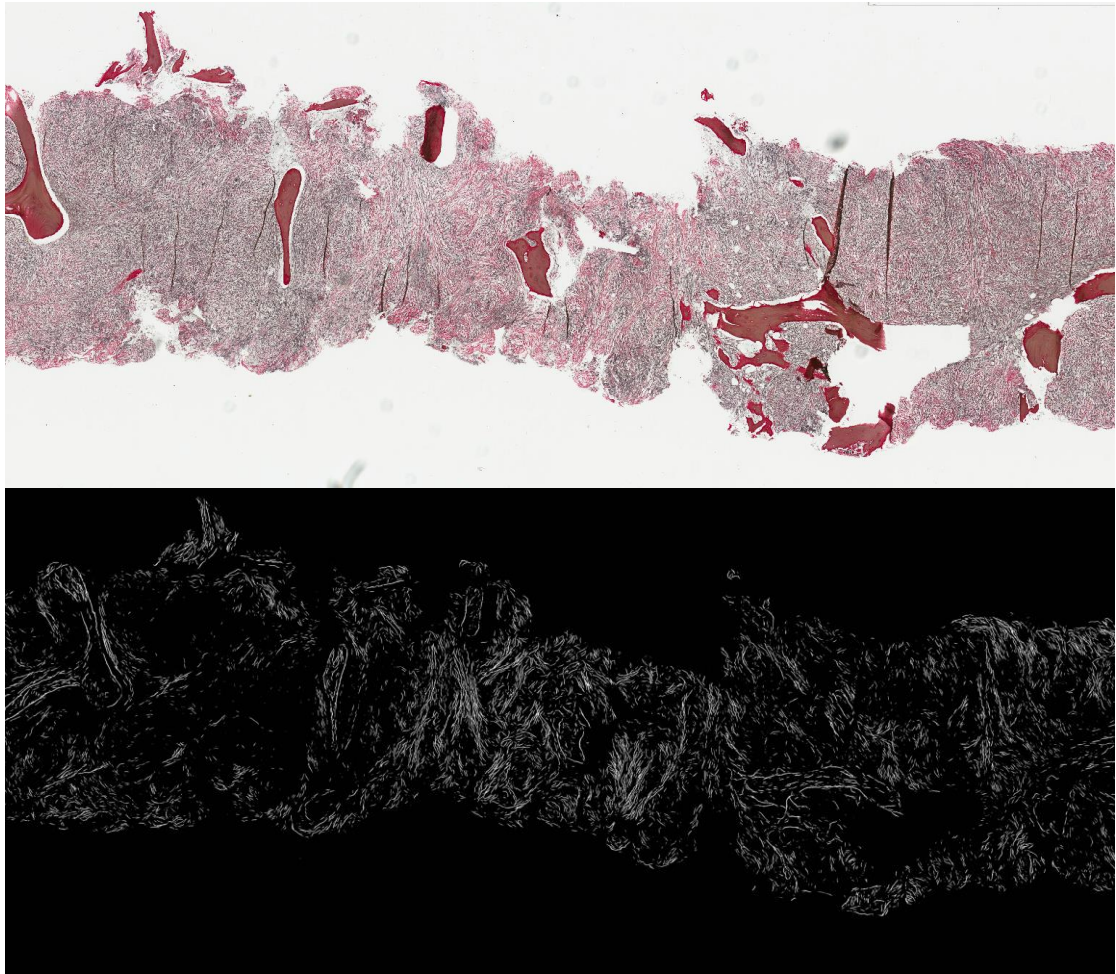
Priedai

1S lentelė. Kolageno struktūros požymiai.

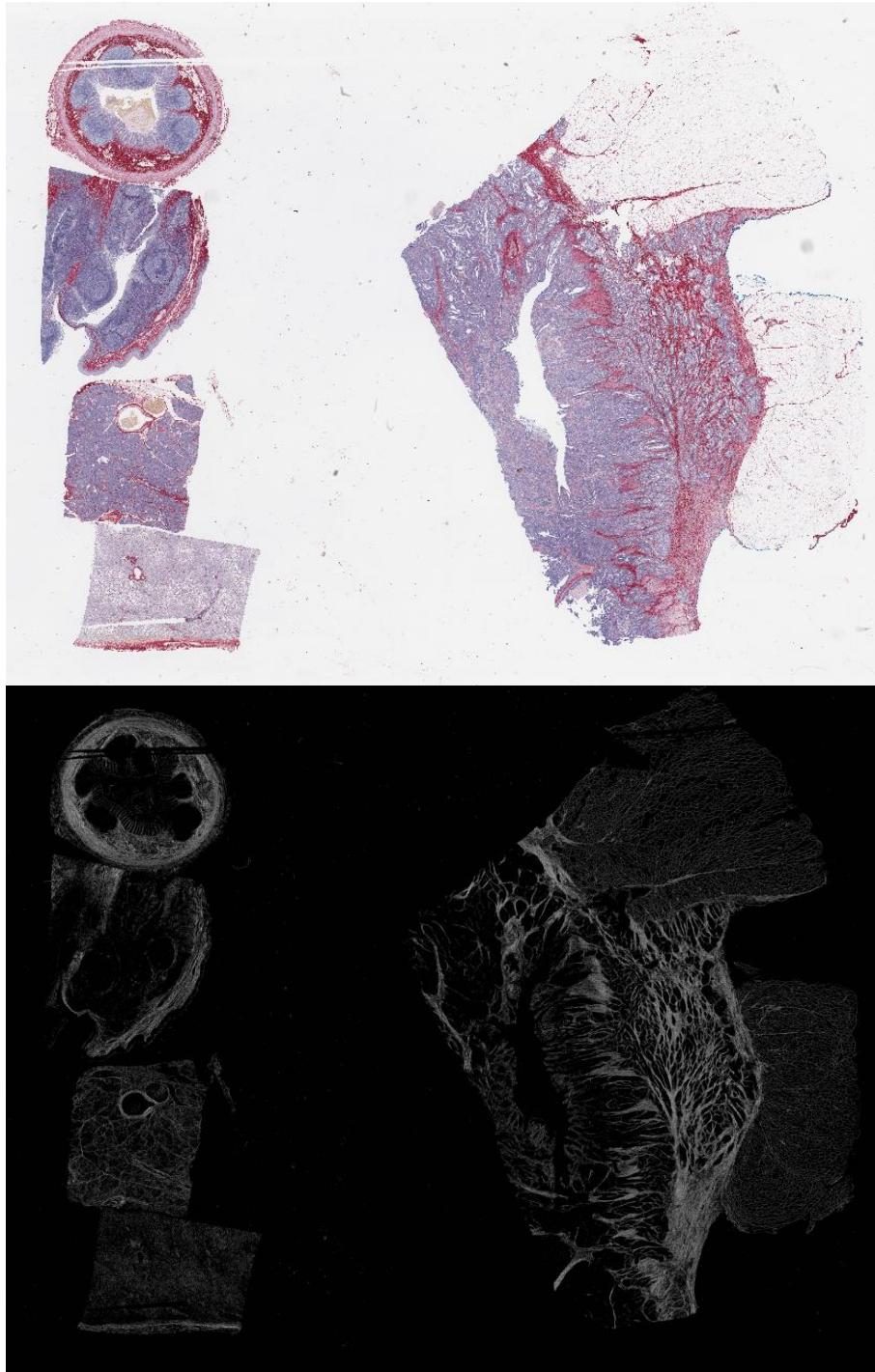
Gijų geometrija		
FW	Plotis (Fiber width)	
FL	Ilgis (Fiber length)	
FP	Vidurio linija (Fiber path)	
FS	Tiesumas (Fiber straightness)	= FL / FP
FD2	Tankis (Fiber density), n=2	vidutinis atstumas iki n kaimynų
FD4	Tankis, n=4	
FD8	Tankis, n=8	
FD16	Tankis, n=16	
Gijų kryptis		
stdMag	Krypties stiprumo variacija	
meanMag	Krypties stiprumo vidurkis	
medianMag	Krypties stiprumo mediana	
stdAng	Krypties kampo variacija	
Kolageno struktūros fraktalinės charakteristikos		
frdMag	Fraktalinė dimensija	
frdAng		
lacMag	Lakunarumas	
lacAng		
Kolageno struktūros tekstūros parametrai		
enrMag	Energija	
enrAng		
conMag	Kontrastas	
conAng		
homMag	Homogeniškumas	
homAng		
entAng	Entropija	
entMag		



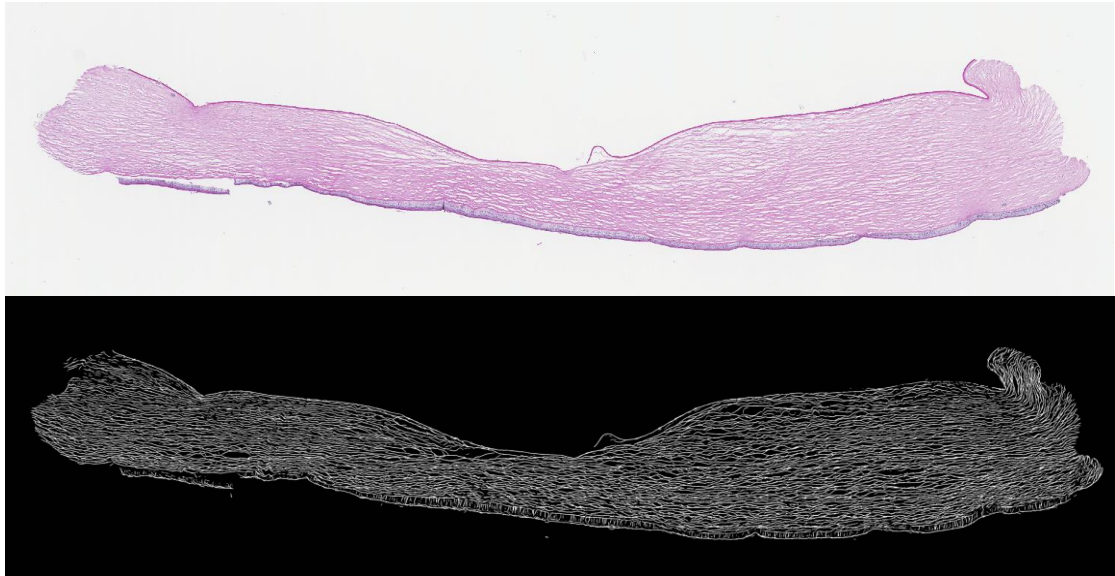
1S pav. Neuroninio tinklo atpažinta kolageno struktūra.



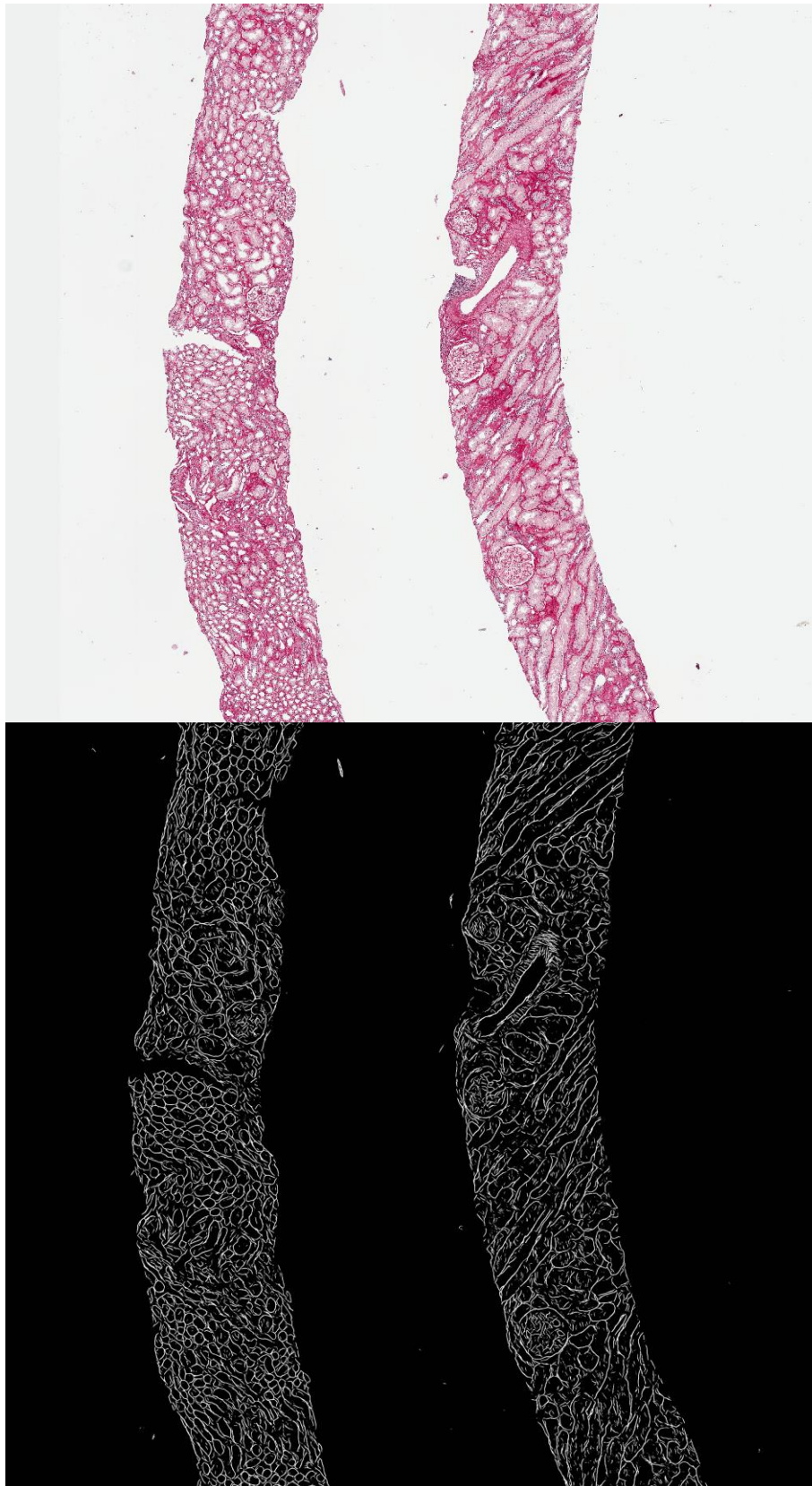
2S pav. limfoma, kaulų čiulpų audinyje. Preparatas dažytas GSPS metodu (viršuje) ir kolageno struktūra (apačioje)



3S pav. storosios žarnos vėžio preparatas dažytas „Sirius-Red“ metodu (viršuje) ir kolageno struktūra (apačioje)



4S pav. Sveikos ragenos preparatas dažytas PAS metodu (viršuje) ir kolageno struktūra (apačioje)



5S pav. Inksto biopsija nudažyta „Sirius-Red“ metodu (viršuje) ir kolageno struktūra (apačioje)