



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



INFORMATIKOS INŽINERIJA (T007)

MODELIS NAVIKŲ GENETINIŲ
POKYČIŲ IR BIOŽYMENŲ RAIŠKOS
DUOMENŲ INTEGRAVIMUI

Mindaugas Morkūnas

2019 m. spalio

Mokslinė ataskaita DMSTI-DS-T007-19-13

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

Vilnius LT-08412

www.mii.lt

Santrauka

Navikų mikroaplinka subtiliai kinta progresuojant patologijai. Šių pokyčių atpažinimas ir detalus kiekybinis charakterizavimas yra itin svarbus pacientų būklės ir ligos eigos prognozavimui, bei biologinių procesų skatinančių navikų vystymąsi supratimui.

Šiame tyrime naudojome anksčiau mūsų pasiūlytą neuroninio tinklo modelį atpažinti krūties vėžio navikų kolageno karkaso struktūrą šviesinės mikroskopijos vaizduose. Modelis buvo adaptuotas darbui su pilno pjūvio skaitmeniniais patologijos vaizdais. Neuroniniu tinklu nuspėtuose kolageno gijų segmentavimo žemėlapiuose identifikuoti objektai (pilnai sujungti komponentai) formuoja pamatinį navikų mikroaplinkos komponentą - tinklo pavidalo kolageno karkasą. Šiai struktūrai aprašyti buvo išgauta daugiamatė (N = 27) požymių erdvė. Navikų mikroaplinka buvo tiriama išsaugant erdvinę informaciją - atskleista kolageno tinklo požymių svarba skirtingai nuo navikinių ląstelių populiacijų nutolusiuose zonose. Įvertinta kolageno tinklo požymių asociacija su tradiciniais patologijos rodikliais.

Tyrimu nustatyta patikima kolageno gijų orientacijos požymių asociaciją su navikų diferenciacijos laipsniu G, morfometrijos požymių asociaciją su navikų TNM klasifikacijos T kategorija, kolageno tankio asociaciją su TNM klasifikacijos N kategorija, tekstūros požymių asociaciją su molekulinio krūties navikų tipu. Nustatyti kolageno tinklo gijų tekstūros, tankio ir krypties požymiai leidžiantys patikimai prognozuoti krūties vėžio pacientų ligos eigą. Pasiūlyta metodika kolageno gijų segmentavimo žemėlapiuose aptikti žinomus svarbius biologinius kolageno parašus - kolageno sutankėjimą greta naviko ląstelių populiacijų, statmenai į naviko ląsteles orientuotas kolageno gijas, lygiagrečiai naviko ląstelėms orientuotas kolageno gijas.

Remiantis tyrimo rezultatais kolageno struktūros požymiai galės būti naujais būdais apjungiami į sudėtingesnius semantinius navikų mikroaplinkos požymius - biologinius parašus, o tyrime pasiūlyta metodika leis automatinį jų aptikimą navikų pilno pjūvio skaitmeniniuose vaizduose.

Reikšminiai žodžiai: navikų mikroaplinka, biologinis parašas, šviesaus lauko mikroskopija, gilaus mokymo neuroninis tinklas.

Turinys

1	Įvadas	4
2	Ekperimentai ir rezultatai	5
2.1	Neuroninis tinklas	5
2.2	Kolageno struktūros požymių erdvė	6
2.3	Pacientai	8
2.4	Kolageno struktūros požymių sąsaja su tradiciniais patologijos rodikliais....	9
2.5	Prognostinė kolageno struktūros požymių reikšmė krūties vėžio pacientų išgyvenamumui	10
2.6	Semantiniai vaizdo požymiai kolageno struktūrai nusakyti	11
3	Literatūra	12

1 Įvadas

Nepaisant įvairių su vėžiu susijusių gairių vis dar sudėtinga pagerinti tam tikrų vėžio tipų pacientų ilgalaikę prognozę. Vis daugiau tyrimų kreipia dėmesį į navikų mikroaplinkos vaidmenį vėžio progresijoje siekdami pagerinti pacientų diagnozavimą, gydymą ir priežiūrą. Yra žinoma, kad pagrindinis navikų mikroaplinkos komponentas - kolagenas progresuojant navikui kinta su tam tikrais dėsniniais. Pakitęs kolagenas aptinkamas praktiškai visuose vėžio tipuose. Šie pokyčiai išreikškiami skirtingais biologiniais parašais (*angl.* tumor associated collagen signatures, TACS): TACS-1 - kolagenas sutankėja greta navikinių ląstelių populiacijų, TACS-2 - kolageno gijos orientuojamos lygiagrečiai naviko ląstelių populiacijos perimetrui ir TACS-3 - kolageno gijos organizuojamos statmenai į naviko ląstelių populiacijos perimetrą [1,2].

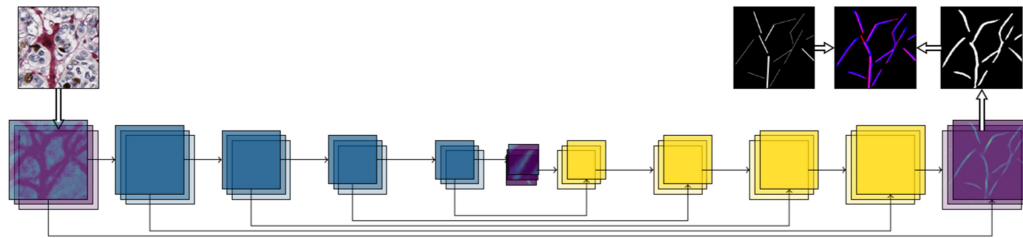
Kolageno struktūros vizualizavimui navikų (ir bendrai visų žmogaus audinių) preparatuose naudojami labai įvairūs metodai - antros harmonikos generacijos, poliarizuotos šviesos ir skystų kristalų poliarizuotos šviesos, elektroninė ir skenuojanti elektroninė, atominės jėgos mikroskopijos. Dažnu atveju tai brangi ir sudėtinga įranga reikalaujanti ypatingo audinių paruošimo ir ribojama tiriamo audinio dydžio [3-6].

Taigi, turi būti sukurti greiti, pigūs ir efektyvūs metodai leidžiantys tirti biologinę informaciją slypinčią kolageno struktūroje. Todėl tyrėme kolageno karkaso segmentavimą diagnostiniuose pilno pjūvio šviesinės mikroskopijos patologijos vaizduose (PPV) naudodami neuroninį tinklą (NN) ir pademonstravome išgautą kolageno struktūros požymių svarbą pacientų prognozei ir navikuose vykstančių procesų supratimui.

2 Eksperimentai ir rezultatai

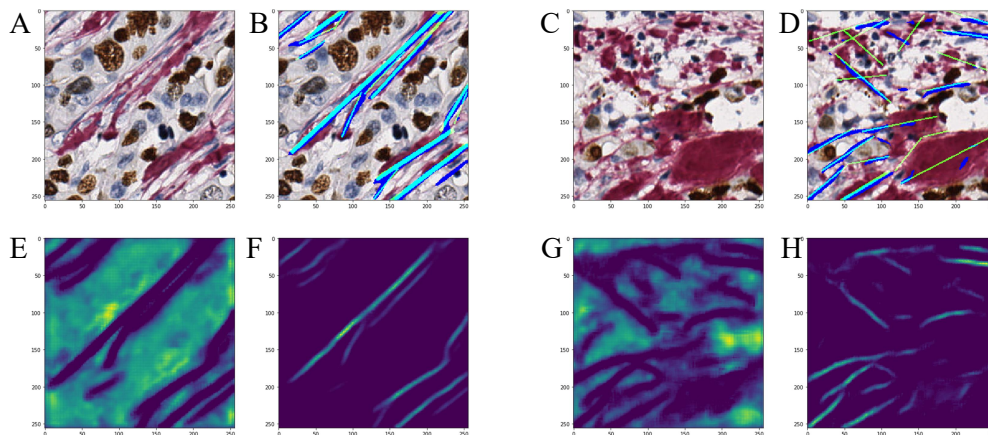
2.1 Neuroninis tinklas

Šiame tyrime naudojome anksčiau mūsų pasiūlytą neuroninio tinklo modelį (žr. 1 pav.) atpažinti krūties vėžio navikų kolageno karkaso struktūrą šviesinės mikroskopijos vaizduose.



1 pav. „U-net“ architektūra pagrįsta neuroninio tinklo struktūra.

Apmokymo duomenų rinkinys buvo rankiniu būdu anotuotas taip, kad neuroninis tinklas reaguodamas į įvesties vaizdą mokintūsi išryškinti skaidulines struktūras vaizdo vietose su akivaizdžiai išreikšta anizotropija (žr. 2 pav.).



2 pav. NN atsakas į įvesties vaizdus. A ir C - originalūs įvesties vaizdai, B ir D - ant originaliųjų vaizdų pateiktos anotacijos (žalia spalva) ir modelio predikcijos (mėlyna spalva). Žydra spalva pažymėtos vietos kur anotacijos persidengia su predikcijom. E, F, G ir H - paskutinio NN sluoksnio (prieš įvesties sluoksnį) atsako į įvesties vaizdus žemėlapiai (*angl.* kernel outputs). 2 iš 32 atsako žemėlapiai su labiausiai išreikštais fono (E ir G) ir ieškomo objekto (F ir H) signalais.

Neuroninis tinklas paslėptuose sluoksniuose išmoksta atpažinti tiek anizotropiškas, tiek orientacijos požiūriu neinformatyvias vaizdo vietas. Neuroninio tinklo išvesties

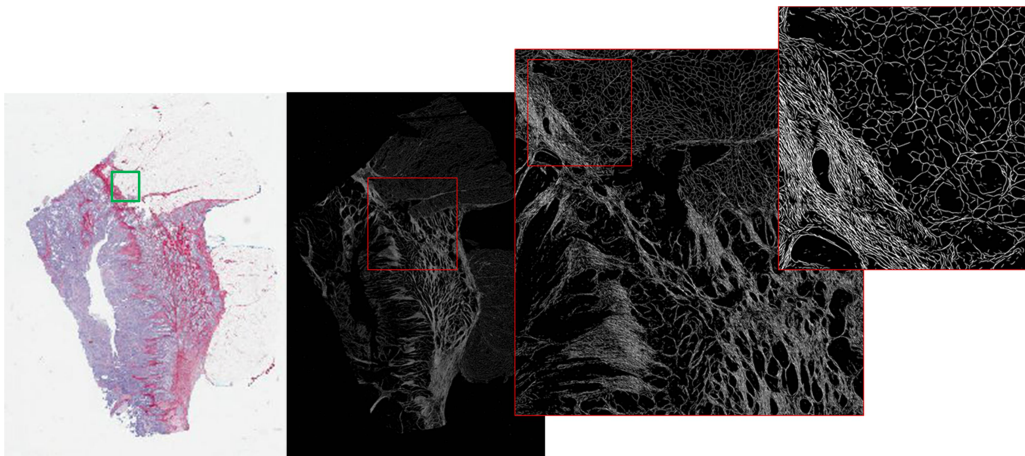
sluoksnyje kolageno gijos gimsta anizotropiškose vaizdo vietose ir miršta ten kur vaizdo turinys nuo orientacijos nepriklauso.

Modelis buvo adaptuotas darbui su PPV vaizdais, dalinant juos į 256 x 256 pikselių dydžio persidengiančias iškarpas, kurios po vieną buvo paduodamos į neuroninį tinklą nuspėti kolageno segmentavimo žemėlapiui. 256 x 256 pikselių dydžio tinklo išvesties sluoksnyje gautas vaizdas buvo binarizuojamas taikant 0,5 slenkstinę reikšmę (tinklas išvestyje atiduoda tikimybinį žemėlapi, kurio kiekvienas pikselis gauna vertę tarp 0 ir 1). Apdoroti neuroninio tinklo tikimybiniai žemėlapiai (256 x 256 pikselių dydžio) buvo apjungiami į pilno pjūvio vaizdą atitinkantį kolageno segmentavimo žemėlapi, sukljuojant iškarpas jų persidengimo vietose.

2.2 Kolageno struktūros požymių erdvė

Grubus kolageno karkaso motyvų supratimas buvo gautas ekspertui anotuojant originalius vaizdus. Jokių alternatyvių segmentavimo rezultatų, ar kitokio etalono neturėjome, todėl neuroninis tinklas buvo mokomas atkurti žmogaus regimąjį suvokimą, o modelio efektyvumas (tikslumas segmentuojant naujus vaizdus) buvo vertinamas peržiūrint segmentavimo rezultatus „plika akimi“ (žr. 2 pav.).

Neuroniniu tinklu nuspėtuose kolageno gijų segmentavimo žemėlapuose identifikuoti objektai formuoja sudėtingą struktūrą (žr. 3 pav.), kuri atitinka pamatinį navikų mikroaplinkos komponentą tinklo pavidalo kolageno karkasą.



3. pav. PPV vaizde išsegmentuotas kolageno karkasas. Originalus paveikslėlis yra 57767 x 45287 pikselių (7,31Gb) dydžio, gautas 20x didinimu.

Kolageno karkasui būdinga tam tikra organizacinių lygių hierarchija. Pats primityviausias šios struktūros organizavimo lygis gali būti nusakomas objektų skaičiavimu grįstais principais, tokiais kaip tankis. Tyrime kolageno struktūros tankiui aprašyti naudojome tokius požymius kaip gijų galų tankis, vidutinis atstumas tarp pasirinkto skaičiaus skirtingų gijų galų ir kitus. Aukštesnį kolageno struktūros organizacinį lygį aprašėme morfometriniiais požymiais - tokiais kaip atstumas tarp tos pačios gijos galų (ilgis), gijos vidurio linijos (skeleto) ilgis, plotis. Orientacijos ir galiausiai lokaliai tekstūrą aprašantys požymiai yra aukščiausių organizacijos lygį nusakantys požymiai esantys arčiausiai semantinio požymių lygmenis.

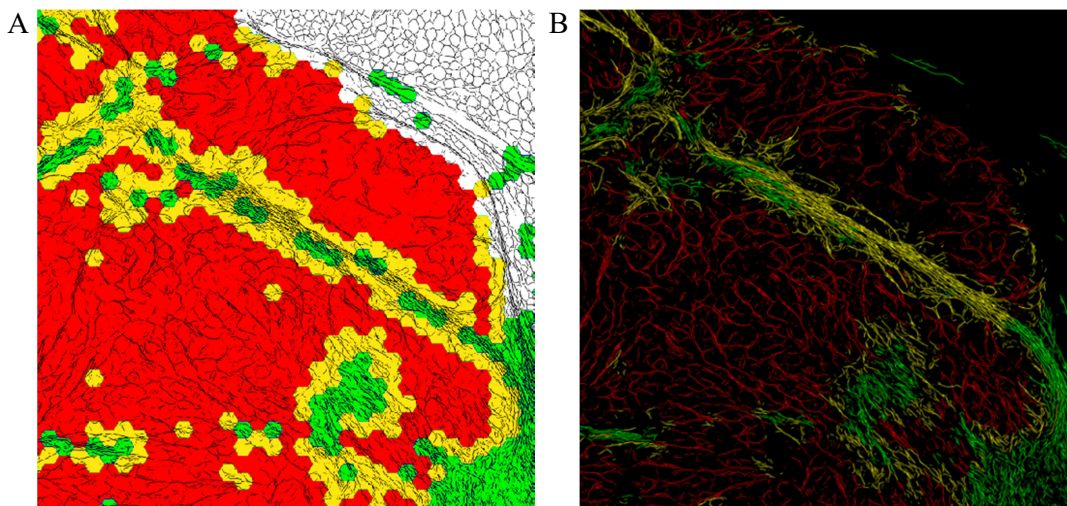
1 lentelė. Kolageno struktūrą aprašančių požymių erdvė.

Eil. nr.	Trumpinys	Požymis	Grupė
1	FL	gijos ilgis, storis, gijos vidurio linijos ilgis, gijos tiesumas (FL / FP)	morfometrija
2	FP		
3	FW		
4	FS		
5	FD	gijų sudarančių pikselių suma, gijos galų skaičius, atstumų tarp visų galinių taškų ploto vienetu vidutinė vertė ir variacija	tankis
6	nENDP		
7	meanEFD		
8	medianEFD		
9	stdEFD		
10	dirmeanANG	dominuojančios gijų orientacijos kampas ir krypties stiprumas, jų variacija ir standartinis nuokrypis	orientacija
11	cirvarANG		
12	cirstdANG		
13	meanMAG		
14	stdMAG		
15	energy		tekstūra
16	contrast		
17	correlation		
18	sum_of_squares_variance		
19	inverse_difference_moment		
20	sum_average		
21	sum_variance		
22	sum_entropy		
23	entropy		
24	difference_variance		
25	difference_entropy		
26	Information_measure_of_correlation_1		
27	Information_measure_of_correlation_2		

Tyrime užsibrėžėme tikslą požymius išgauti ir jų reikšmę įvertinti erdvinės informacijos kontekste, todėl vaizdo erdvė buvo sudalinta į režius pagal atstumą d iki artimiausios naviko ląstelių populiacijos. Visas tiriamas audinys buvo suklasifikuotas į tris klases. 4 paveiksle geltona spalva pavaizduota navikinio ir jungiamojo audinio sąveikos zona ($d = 0$, t.y. tiriamo audinio vietos, kuriose navikinis audinys maišosi su

jungiamuoju audiniu). Raudona spalva pavaizduotas navikinis audinys, sudarytas didžiaja dalimi iš naviko ląstelių ($d > 0$, atspindi kaip giliai navikiniame audinyje yra tiriami audinio dalis). Žalia spalva pavaizduotas jungiamasis audinys ($d < 0$, atspindi kaip giliai jungiamajame audinyje yra tiriami audinio dalis).

Požymiai buvo išgaunami ir agreguojami į paciento lygmenį dviem būdais - požymiai išgaunami lokaliai naudojant šešiakampių gardeles, arba požymiai buvo išgaunami kiekvienam kolageno tinklo elementui (pilnai sujungtam komponentui). Į paciento lygmenį požymiai buvo agreguojami kur įmanoma skaičiuojant vidurkį, medianą ir standartinę nuokrypį.



4. pav. Erdvinis kolageno požymių kontekstas. A - požymiai išgaunami ir agreguojami šešiakampėse gardelėse; gardelės klasifikuotos pagal atstumą iki artimiausios naviko ląstelių populiacijos. B - atitinkamai klasifikuotos kolageno gijos (kaip atskiri objektai).

2.3 Pacientai

Tyrimui buvo gautas Lietuvos Bioetikos komiteto leidimas. Tyrimui naudojome 101 krūties vėžio pacientės histologinių preparatų PPV vaizdus. Šių pacientų klinikiniai-patologijos rodikliai pateikiami 2 lentelėje. Pagrindiniai iš jų: T – atspindi pirminio naviko dydį, N – sritinių limfmazgių būklę, M – atokiąsias metastazes. Šių simbolių reikšmės įvairių lokalizacijų atžvilgiu šiek tiek skiriasi, tačiau pagrindiniai jų principai išlieka. T1-4 rodo naviko dydį centimetrais (T1 – mažiausias, T4 – didžiausias). Simboliai N0-3 rodo limfmazgių, kuriuose aptikta metastazių, skaičių. M0-2 atokiosios metastazės viename ar keliuose organuose. G1-4 žymi naviko

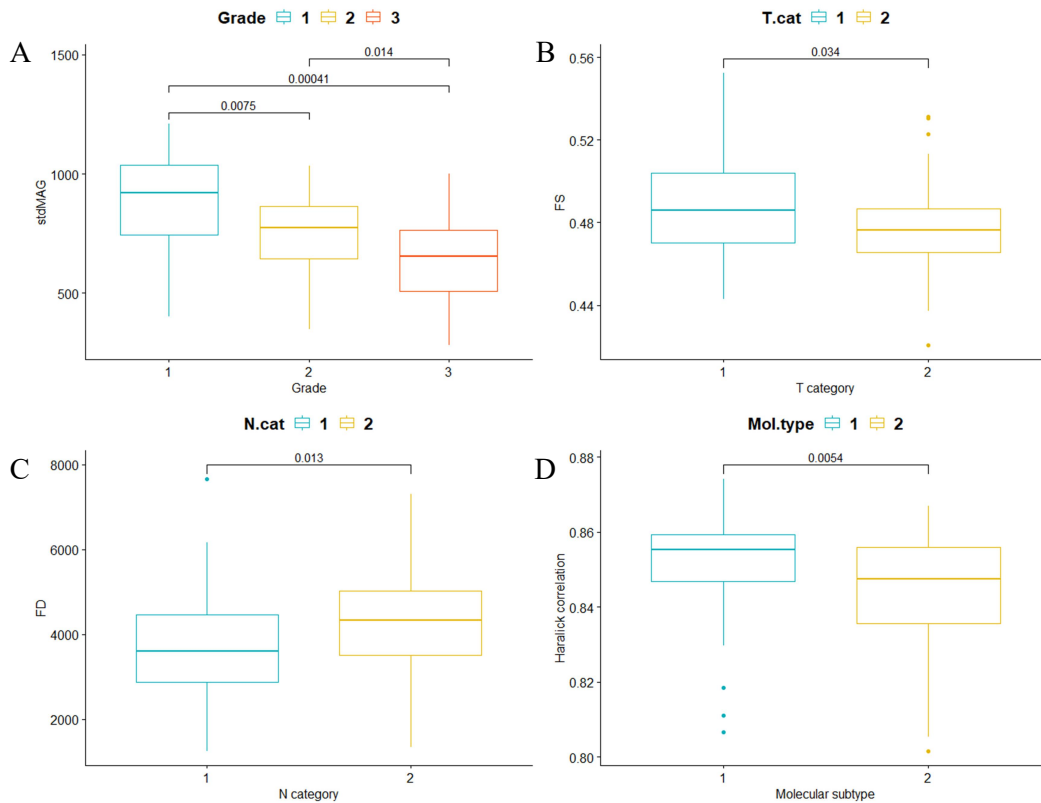
diferenciacijos laipsnį. Kuo naviko struktūra labiau skiriasi nuo sveiko audinio t. y. mažiau diferencijuota, tuo ligos eiga agresyvesnė. Aukšto laipsnio, t.y. geros diferenciacijos navikas žymimas simboliu G1, mažos diferenciacijos – G3–4. Krūties navikai dar papildomai klasifikuojami į molekulinis-histologinius tipus. Luminalinis A tipas, Luminalinis B tipas (HER2 neigiamas potipis), Luminalinis B tipas (HER2 teigiamas potipis), HER2 tipas, Bazalinis tipas.

2 lentelė. Klinikiniai-patologiniai pacientų rodikliai.

T kategorija	T1	T2	T3	T4	VISO
	55	46	0	0	101
N kategorija	N0	N1	N2	N3	
	54	35	9	3	101
	N0	N1-2-3			
	54	47			101
G laipsnis	G1	G2	G3	G4	
	23	47	31	0	101
	G1-2		G3		
	70		31		101
Molekulinis tipas	Lum.A		Lum.B HER2+	Lum.B HER2-	
	56		16	29	101
	Lum.A		Lum.B		
	56		45		101
Amžius	<59m.		>59m.		
	53		48		101

2.4 Kolageno struktūros požymių sąsaja su tradiciniais patologijos rodikliais

ANOVA analize nustatėme patikimą kolageno gijų orientacijos požymių asociaciją su navikų diferenciacijos laipsniu G, morfometrijos požymių asociaciją su navikų TNM klasifikacijos T kategorija, kolageno tankio asociaciją su TNM klasifikacijos N kategorija, tekstūros požymių asociaciją su molekulinis krūties navikų tipu (žr. 5 pav.).

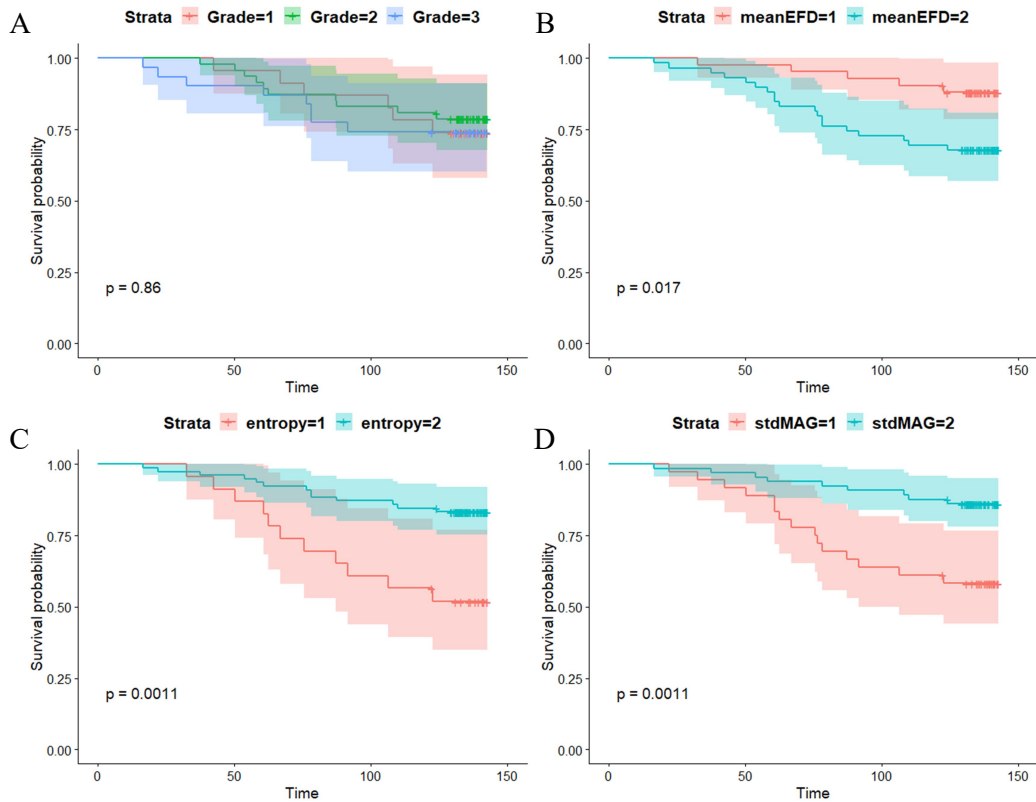


5 pav. ANOVA analizė su *post hoc* Tukey HSD (ganėtinai statistiškai reikšmingo skirtumo) testu. A - kolageno gijų orientacijos stiprumas navikiniame audinio komponente patikimai skiriasi tarp navikų su skirtingu diferenciacijos laipsniu (G); B - kolageno gijų tiesumas navikiniame audinio komponente patikimai skiriasi tarp navikų su skirtingomis T kategorijomis; C - kolageno gijų tankis patikimai skiriasi tarp skirtingų N kategorijų navikų; D - Haraliko tekstūros koreliacijos koeficientas patikimai skiriasi tarp skirtingų molekulinų tipų navikų.

2.5 Prognostinė kolageno struktūros požymių reikšmė krūties vėžio pacientų išgyvenamumui

Pacientės po operacijos buvo stebimos 17-142 mėnesius, todėl kolageno struktūros požymių reikšmingumą buvo galima įvertinti išgyvenamumo analizę taikant Kaplano ir Mejerio metodą (grupėms palyginti naudotas „log-rank“ testas). Iš esmės, tyrimui pasirinkta taip vadinamų geros prognozės pacientų imtis (vidutinė gyvenimo trukmė po operacijos siekė 121 mėnesį, tačiau vis tiek per stebėjimo laikotarpį registruoti 24 įvykiai). 6A paveiksle matome, jog gydytojo patologo nustatomas G laipsnis neleidžia prognozuoti šių pacientų išgyvenamumo, tačiau kolageno struktūros požymiams galima rasti slenkstines reikšmes, kurios leidžia patikimai stratifikuoti pacientus į

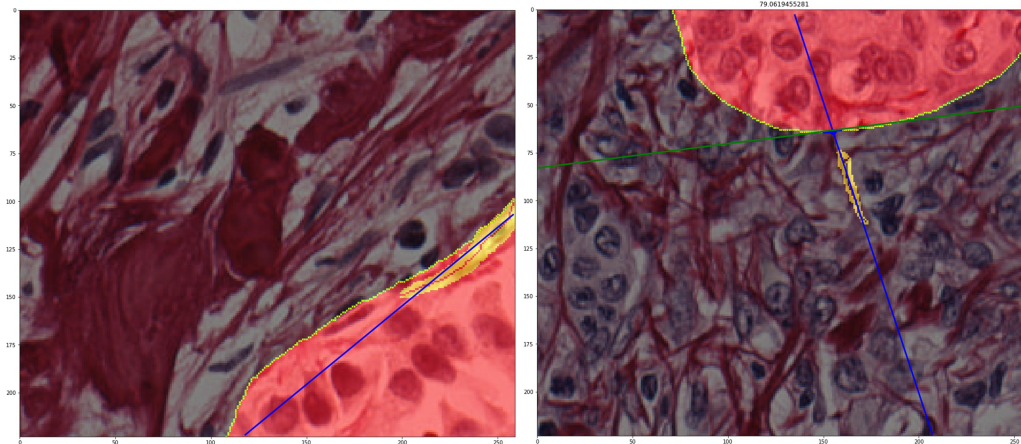
prognostines grupes. Atstumų tarp kolageno gijų galų vidurkis naviko sąveikos zonoje (žr. 6B pav.), kolageno gijų entropija navikiniame komponente (žr. 6C pav.) ir kolageno gijų orientacijos stiprumo variacija jungiamojo audinio komponente (žr. 6D pav.) leidžia patikimai prognozuoti krūties vėžio pacientų ligos eigą.



6 pav. Pacientų išgyvenamumo analizė. A - gydytojo patologo nustatytas G laipsnis neleidžia patikimai prognozuoti pacientų išgyvenamumo; B, C ir D - patikimas ir reikšmingas pacientų stratifikavimas į geros ir blogos prognozių grupes.

2.6 Semantiniai vaizdo požymiai kolageno struktūrai nusakyti

Semantinis požymių lygmuo kolageno struktūrai nusakyti šiuo metu apsiriboja trimis požymiais, taip vadinamais biologiniais kolageno parašais ir šiame tyrime buvo nagrinėtas tik dalinai. Pasiūlyta metodika kolageno gijų segmentavimo žemėlapiuose aptikti du iš trijų biologinių kolageno parašų - matuoti kolageno gijų orientaciją naviko ląstelių atžvilgiu. Tokiu būdu galima aptikti statmenai ir lygiagrečiai į naviko ląstelių populiacijas orientuotas kolageno gijas (žr. 7 pav.).



7 pav. Kolageno gijų orientacija link artimiausios navikinių ląstelių populiacijos gali būti išmatuota automatiškai visame PPV vaizde.

Tikimės, kad remiantis šio tyrimo rezultatais kolageno struktūros požymiai galės būti naujais būdais apjungiami į sudėtingesnius semantinius navikų mikroaplinkos požymius, o tyrime pasiūlyta metodika leis automatinį jų aptikimą navikų pilno pjūvio skaitmeniniuose vaizduose.

3 Literatūra

1. 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.
2. Conklin, M. W. et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. *The American Journal of Pathology* 178, 1221 – 1232 (2011).
3. Hompland, T et al. (2008). “Second-harmonic generation in collagen as a potential cancer diagnostic parameter”. *J Biomed Opt.* 13(5):054050
4. 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.
5. 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.
6. 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.