



# Skaitmeninių signalų tyrimas ir modeliavimas audinio vėžinėms zonoms aptikti skirtinguose MRI modalumuose.

Doktorantas: Aleskas Vaitulevičius

Vadovas: Doc. dr. Povilas Treigys

Mokslo kryptis: N009 Informatika

Studijų laikotarpis: 2021 – 2025 m.

Studijų metai: 2022/2023



# Pristatymo turinys

- Tyrimo objektas ir tikslas
- Tyrimo uždaviniai
- Visų studijų planas ir jo vykdymo suvestinė
- Ataskaitinių metų darbo planas ir jo įvykdymas
- Visų mokslinių tyrimų ir disertacijos rengimo etapai
- Gauti moksliniai rezultatai
- Kito pusmečio darbo planas



# Tyrimo objektas, tikslas

- Objektas:
  - Magnetinio rezonanso tomografijos dinaminio kontrasto sustiprinimo (angl. Magnetic Resonance Imaging Dynamic Contrast Enhancement (MRI DCE)) modalumo nuotraukų sekos.
  - Mašininio mokymosi ir funkcinių duomenų analizės algoritmai vėžinėms zonoms aptikti prostatoje.
- Tikslas: Nustatyti vėžinės zonos aptikimo galimybes naudojantis standartiniais mašininio mokymosi ir funkcinių duomenų analizės metodais



# Tyrimo uždaviniai

- Sukurti algoritmą, kuris, naudodamas standartinius mašininio mokymosi ir funkcinių duomenų analizės metodus, aptiktų vėžinės zonas prostatoje.
- Aptikti triukšmingą duomenų imties poaibį.
- Empiriniais eksperimentais nustatyti sukurto algoritmo parametrus, su kuriais yra pasiekiami aukščiausi tikslumo matai.

# Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2021/2022)	2	2		1			
II (2022/2023)	2		1				
III (2023/2024)			1		1		
IV (2024/2025)			1		1		
<b>Iš viso:</b>	<b>4</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>		

# Ataskaitinių metų darbo planas ir jo įvykdymas

## Egzaminai

Planas	Įvykdyta	Būklė
Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika	Egzamino pavadinimas: Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika Data: 2022-06-28	Išlaikytas
Daugiamačių duomenų vizualizavimas	Egzamino pavadinimas: Daugiamačių duomenų vizualizavimas Data: 2022-07-08	Išlaikytas

# Ataskaitinių metų darbo planas ir jo įvykdymas

## Dalyvavimas konferencijose

Planas	Įvykdyta	Konferencijos tipas
WSCG 2022, 2022 metų Gegužės 17-20 dienomis, Čekija, Pilzenas	<b>Pristatymo pavadinimas:</b> DCE MRI Modality Investigation for Cancerous Prostate Region Detection: Case Analysis <b>Autoriai:</b> Vaitulevičius, A., Treigys, P., Bernatavičienė, J., Surkant, R., Markevičiūtė, J., Naruseičiūtė, I., Trakymas, M. <b>Konferencijos pavadinimas:</b> WSCG 2022 <b>Data:</b> 2022 metų Gegužės 17-20 dienomis <b>Vieta:</b> Čekija, Pilzenas	Tarptautinė

## Publikacijos

Planas	Įvykdyta	Būklė	Publikacijos tipas
--------	----------	-------	--------------------

# Visų mokslinių tyrimų ir disertacijos rengimo etapai

Darbo pavadinimas	Atlikimo terminai
<p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <p>1.1 Disertacijos tyrimo objekto detalizavimas.</p>	<p>2021 m. spalio – 2022 m. kovas</p>
<p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <ol style="list-style-type: none"><li>1. Prostatos vėžio aptikimo ir agresyvumo vertinimo metodų apžvalga atsižvelgiant į histologinius tyrimus.</li><li>2. Atliktos apžvalgos apibendrinimas ir pateikimas disertacijos analitinės dalies aprašyme.</li></ol>	<p>2021 m. kovas – 2022 m. rugsėjis</p>



# Šio pusmečio moksliniai rezultatai

Empiriškai palyginti šias konfigūracijas:

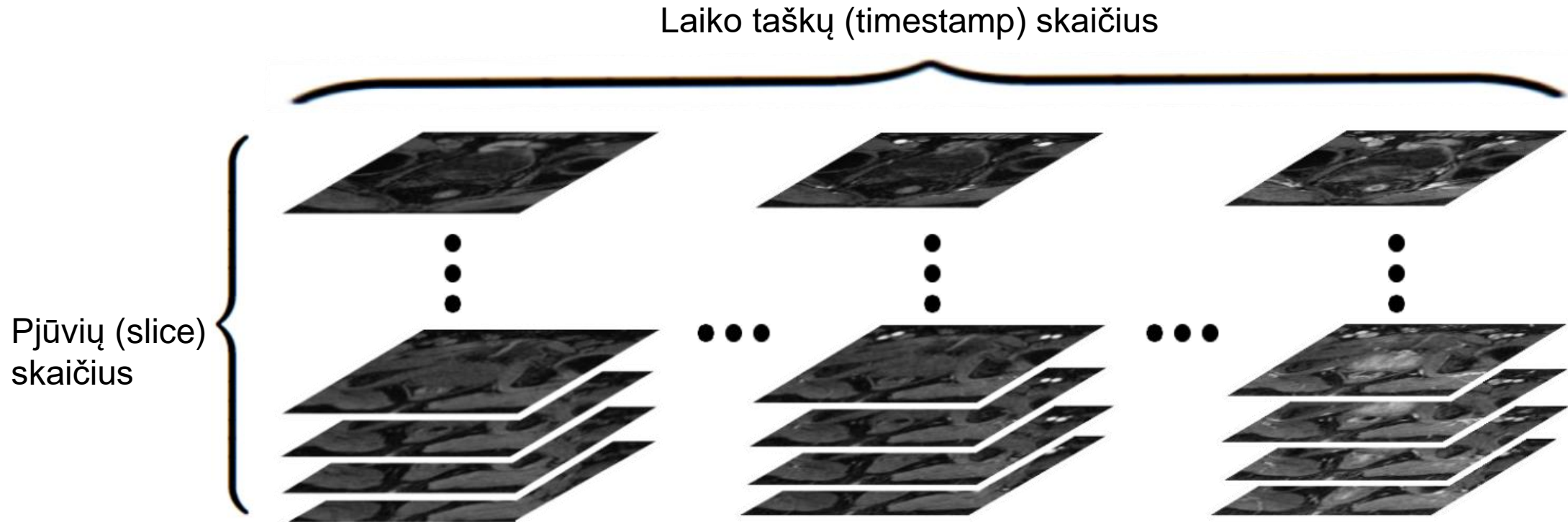
- Fiksuotą SLIC regionų skaičių pjūvyje su proporcingu (fixed number of SLIC regions).
- Registruoti funkciniai duomenys su neregistruotais (registered).
- Multivariaciniai duomenys, kurių požymiai yra ištraukti iš funkcinių duomenų:
  - 100 diskrečių uniformaliai pasiskirsčiusių funkcijos reikšmių
  - integralus gylis
  - modifikuoto diapazono gylis
  - maksimali funkcijos reikšmė
  - argumentas su kuriuo funkcija igyja maksimalią reikšmę.

Su multivariaciniais duomenimis:

- 100 diskrečių uniformaliai pasiskirsčiusių funkcijos reikšmių (only discrete)

Empiriškai nustatyti ar egzistuoja universalūs hiper parametrai XGBoost'ui arba Pilnai sujungtų sluoksnių neuroniniui tinklui.

# Duomenys: DCE-MRI (vieno paciento pavyzdys)



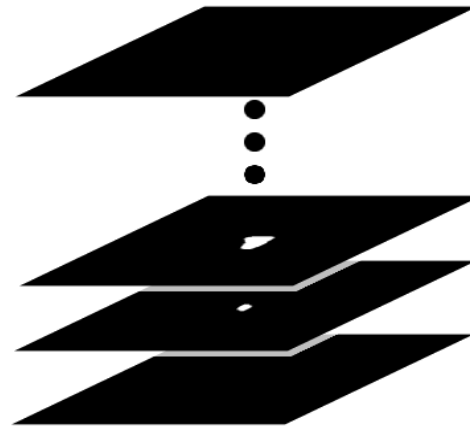
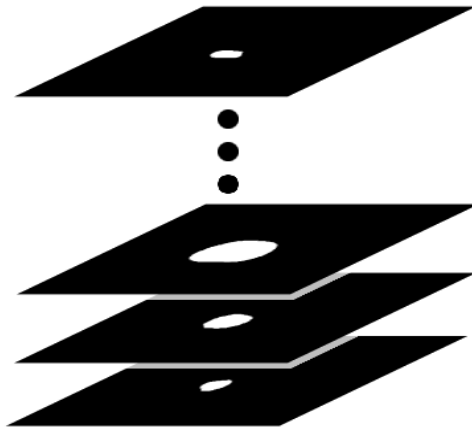
# Duomenys: Kaukės (vieno paciento pavyzdys)

Prostatos regiono  
kaukė

Vėžinio regiono  
kaukė

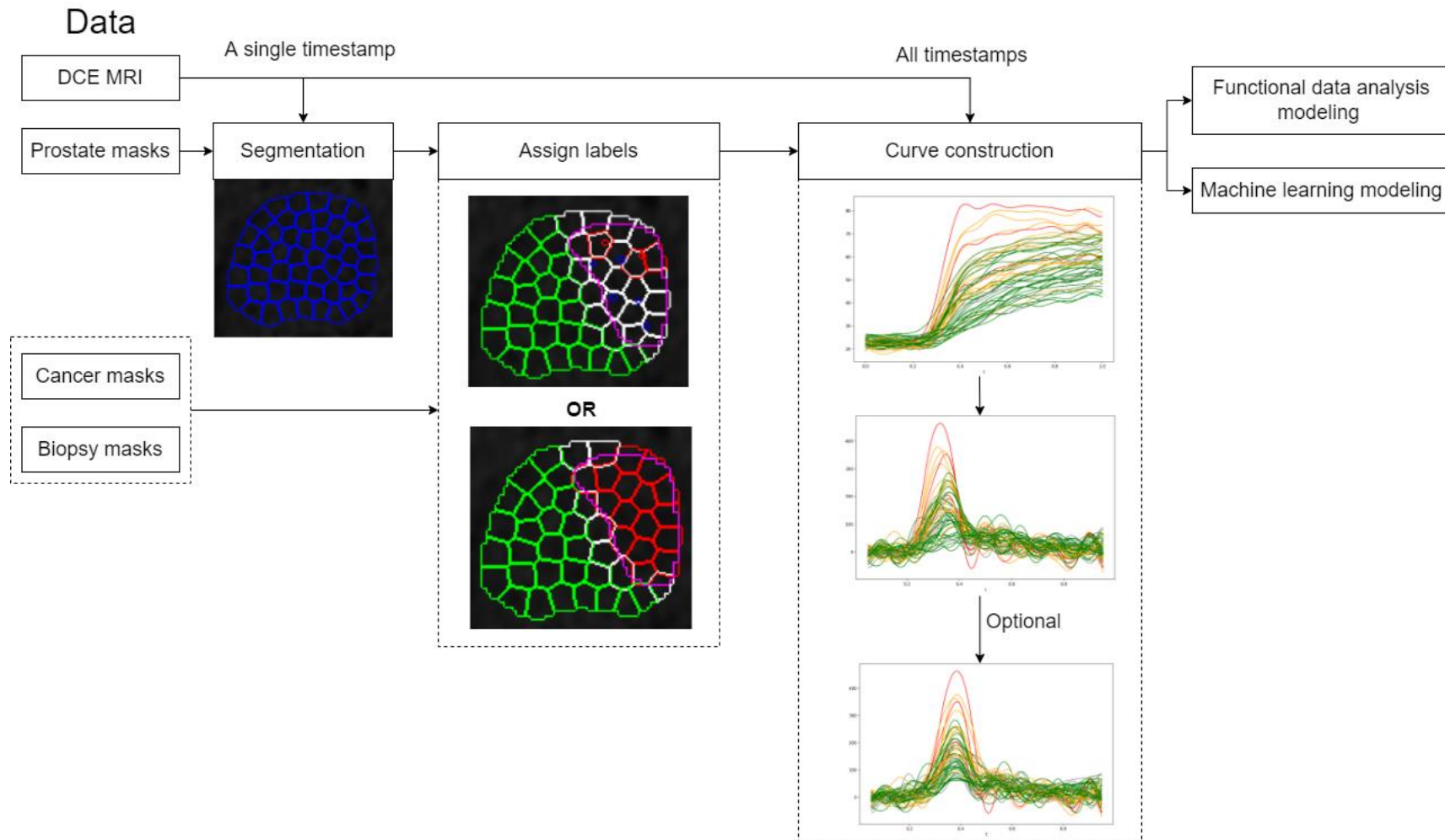
Biopsijų kaukė

Pjūvių (slice)  
skaičius



- Piktybinis vėžys (malignant)
- Neapibrėžtas vėžys (undetermined)
- Nepiktybinis vėžys (benign)

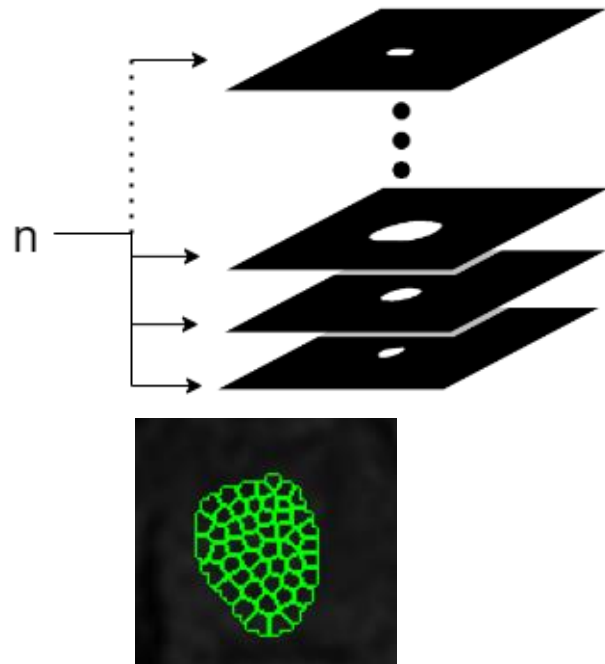
# Veiksmų seka



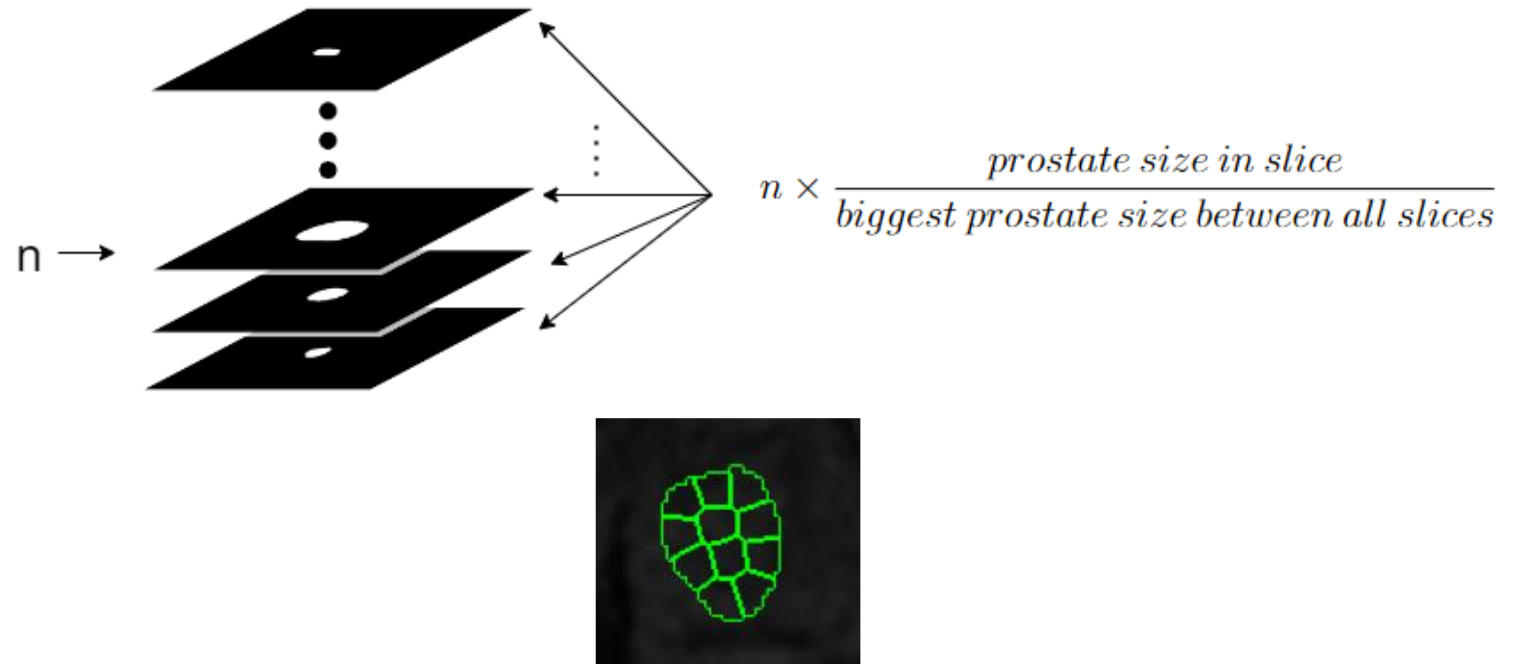
# Segmentavimas

SLIC algoritmu (iteratyvus paprastasis linijinis klasterizavimas).

Fiksuotas SLIC regionų  
Skaičiaus parinkimo būdas

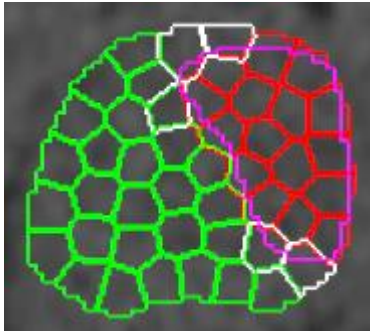


Proporcingas SLIC regionų  
Skaičiaus parinkimo būdas

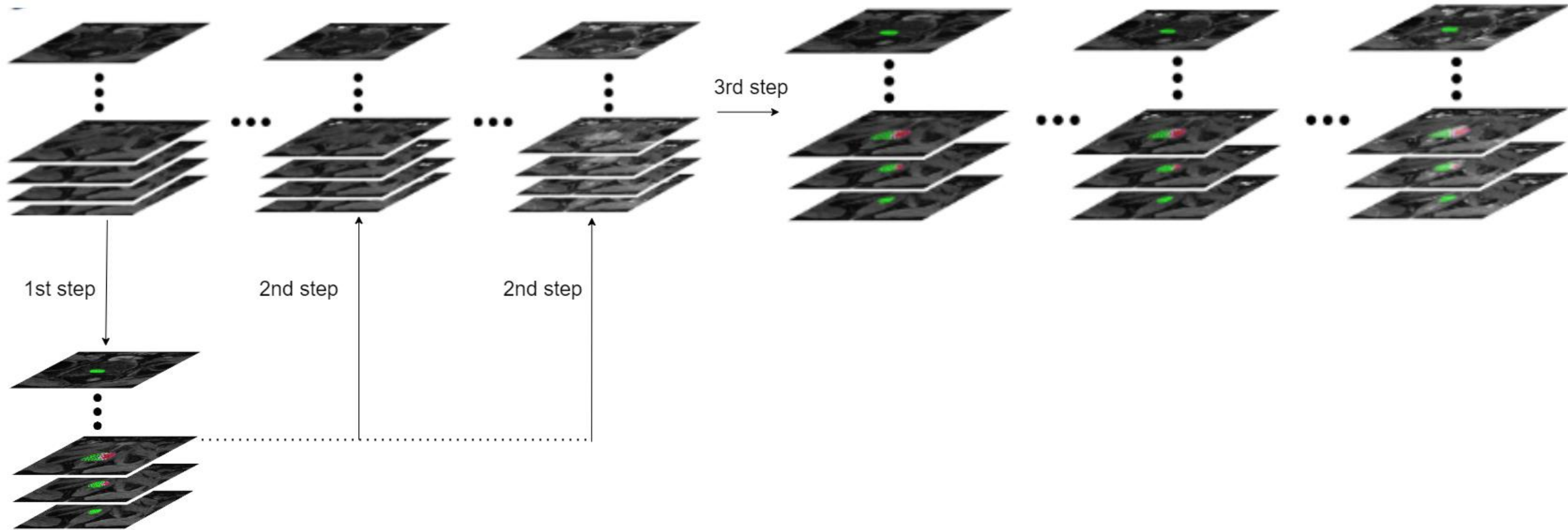


# Klasių priskyrimas

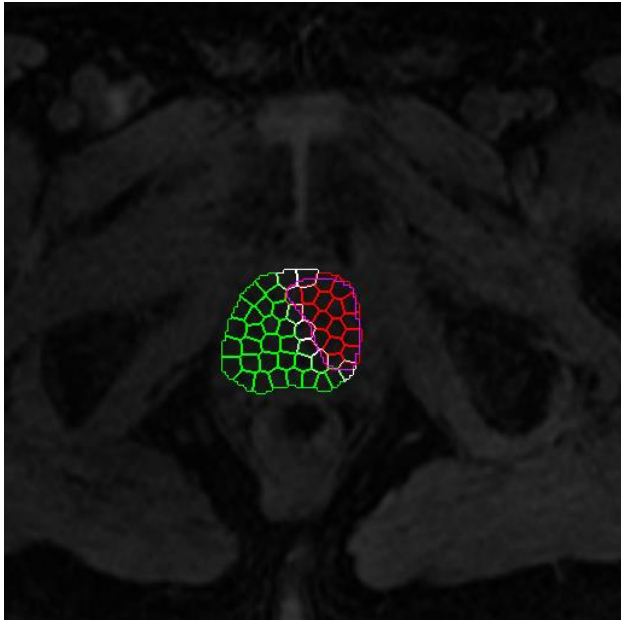
- Su vėžiniu regionu, kuris yra patvirtintas piktybinio vėžio biopsijos, persidengia > 50%
- Neturi persidengimo su vėžiniu regionu, kuris turi persidengimą su piktybinėmis arba neapibrėžtomis biopsijomis.
- Kiti regionai, kurie netenkina kitų grupių sąlygų



# Funkcinių duomenų konstravimas

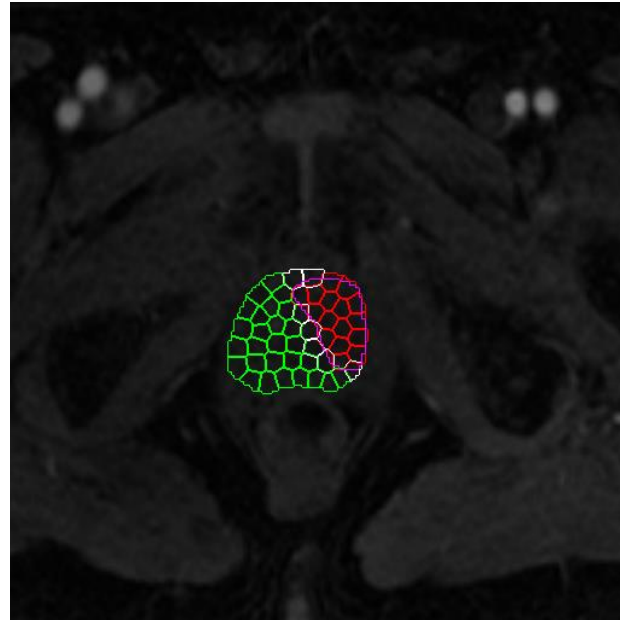


# Funkcinių duomenų konstravimas (pjūvio lygmenyje)



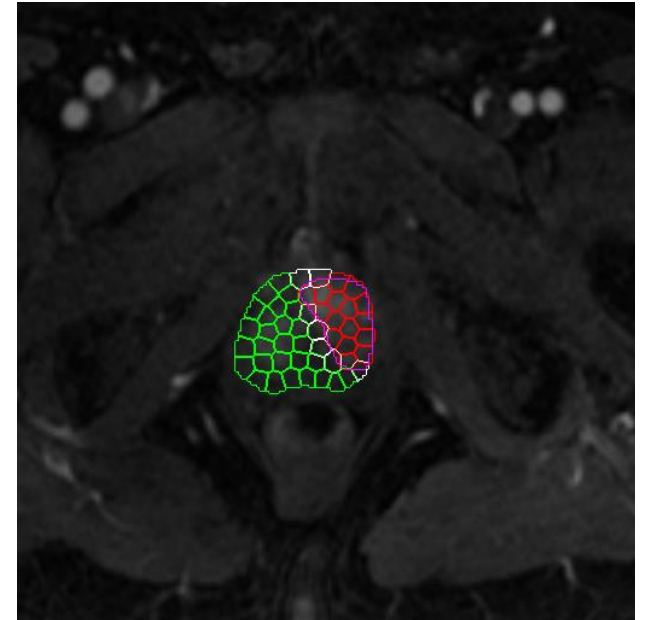
Pirmasis laiko taškas

...



n-tasis laiko taškas

...



(n + m)-tasis laiko taškas



# Funkcinių duomenų konstravimas (laiko eilutė)

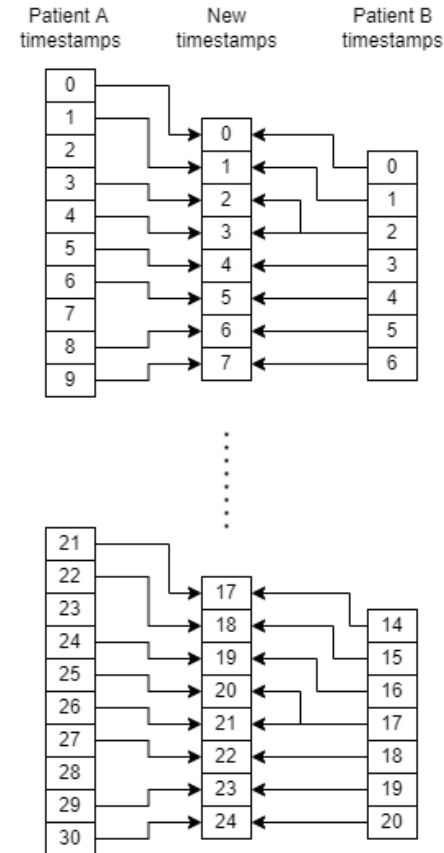
Pjūvis	SLIC regionas	Laiko taškas	Intensyvumų vidurkis
1	1	1	$x_1$
		2	$x_2$
		...	
	2	1	$x_5$
		2	$x_4$
	...		
	...		

# Segmentavimo laiko taško pasirinkimo problema

Pacientai turi skirtingą skaičių laiko taškų. Tad paciento A taškas 5 nėra lygus paciento B laiko taškui 5.

Sprendimas - Pacientų segmentacijų laiko taškų susiejimas (angl. mapping)

Pasirinktas susietas segmentavimo laiko taškas – 5.



# Apmokymas ir hiper parametrų automatinis parinkimas

- Tyrimai atlikti su 13 pacientų duomenimis.
- Apmokymas ir validavimas vykdomas kiekvieno paciento duomenimis atskirai.
- Kiekvienas požymis yra normalizuojamas į interval  $[0, 1]$ .
- Apmokymo duomenų aibė – 70 %, o validavimo – 30%.
- Apmokymui naudojami balansuoti svoriai.
- Modeliams naudotas išankstinis apmokymo stabdymas jei validavimo duomenų F1 įvertis:
  - Nepadidėja po 10 epochų treniruojant pilnai sujungtų sluoksnių neuroninį tinklą.
  - Nepadidėja praiteravus per visus mokymo duomenis 10 kartų treniruojant XGBoost klasifikavimo modelį.
- Pilnai sujungtų sluoksnių neuroniniui tinklui buvo naudojamas Bajeso hiper parametrų automatinio parinkimo algoritmas, o XGBoost klasifikavimo modeliui – Parzeno medžio įvertinimo algoritmą.
- Hiper parametrų automatinis parinkimas yra vykdomas maksimizuojant F1 įvertį.

# Hiper parametrų automatinis parinkimas: XGBoost klasifikavimo modelis

XGBoost klasifikavimo modelio hiper parametrų automatiniam parinkimui treniravimas yra atliekamas 400 kartų su kiekvieno paciento duomenimis atskirai. Treniravimas yra atliekamas minimizuojant logistinės regresijos nuostolių funkciją. Parinktas gradiento augimo medžių (angl. gradient boosted trees) skaičius yra 180. Automatiškai parinkti hiper parametrai yra:

- Stulpelių poaibio santykis konstruojant kiekvieną medį (paieškos erdvė yra intervalas  $[0.4, 0.8]$ ).
- Minimalus nuostolio mažinimas reikalingas atlikti medžio lapo padalinimą (paieškos erdvė yra intervalas  $[0, 1]$ ).
- Maksimalus XGboost ensamblio medžio gylis (paieškos erdvė yra intervalas  $[3, 18]$ ).
- Minimali atskiro atvejo svorių (Hessian) suma reikalinga vaike (paieškos erdvė yra intervalas  $[0, 10]$ ).
- L1 svorių reguliarizacijos terminas (paieškos erdvė yra intervalas  $[0, 1]$ ).



# Hiper parametrų automatinis parinkimas: Pilnai sujungtų sluoksnių neuroninis tinklas

Pilnai sujungtų sluoksnių neuroninio tinklo hiper parametrų automatiniam parinkimui treniravimas yra atliekamas bent 25 kartus su kiekvieno paciento duomenimis atskirai. Kiekvienas treniravimas yra atliekamas su  $\leq 100$  epochomis. Kiekvienas hiper parametrų automatinis parinkimas yra atliekamas 10 kartų. Pilnai sujungtų sluoksnių neuroninio tinklo automatinė architektūros paieškos erdvė yra:

- Įvesties sluoksnis
- Pirmas paslėptasis pilnai sujungtas sluoksnis su ReLU aktyvacijos funkcija ir neuronų skaičiumi lygiu įvesties dimensijoms.
- Paslėptų pilnai sujungtų sluoksnių skaičiaus paieška (neuroninio tinklo gylis) (paieškos erdvė yra intervalas  $[1, 10]$ ). Kiekvieno sluoksnio aktyvacijos funkcija yra ReLU. Neuronų skaičius kiekviename sluoksnyje yra ieškomas imant 32 skaičiaus kartotinius iš intervalo  $[32, 1024]$ .
- Išėjties sluoksnis susideda iš vieno neurono. Šio neurono aktyvacijos funkcija yra automatiškai parenkama. Testuojamos funkcijos yra sigmoidinė, softmax ir ReLU.

# Surinkti tikslumo matai ir jų agregacijos

- Surinkti validacijos duomenų tikslumo matai yra tikslumas (angl. precision), atitikimas (angl. recall), F1 įvertis, balansuotas tikslumas ir specifiškumas vieno paciento ir konfigūracijos lygmenyje. Šios metrikos yra agreguojamos atliekant tokius žingsnius:
  - Pilnai sujungtų sluoksnių neuroninio tinklo hiper parametrų automatinis parinkimas yra atliekamas 10 kartų. Tad apskaičiuojamos gautų įverčių vidurkiai ir standartiniai nuokrypiai.
  - Apskaičiuojamos XGboost klasifikavimo modelio tikslumo matų ir pilnai sujungtų sluoksnių neuroninio tinklo tikslumo matų vidurkių ir standartinio nuokrypių medianos. Medianos apskaičiuojamos lygmeniu:
    - Konfigūracijos
    - Vienos konfigūracijos
- Kiekvienai konfigūracijai yra ištraukiamos paciento, su kuriuo modelis pasiekė aukščiausią balansuotą tikslumą, hiper parametrų automatinio parinkimo rezultatai.

# Surinkti hiper parametrai

Pilnai sujungtų sluoksnių neuroninio tinklo gylis ir neuron skaičius kiekviename sluoksnyje

Pilnai sujungtų sluoksnių neuroninis tinklas

Registruoti	Daugiamačiai požymiai	Segmentavimo metodas	Aktyvacijos funkcija	0	1	2	3	4	5	6	7	8	9	10	
Neregistruoti	Visi požymiai	Proporcingas SLIC	sigmoidinė	864	416										
		Fiksuotas SLIC	sigmoidinė	1024											
	Tik diskretūs taškai	Proporcingas SLIC	sigmoidinė	1024											
		Fiksuotas SLIC	sigmoidinė	1024											
Registruoti	Visi požymiai	Proporcingas SLIC	sigmoidinė	1024	384	320									
		Fiksuotas SLIC	sigmoidinė	1024											
	Tik diskretūs taškai	Proporcingas SLIC	sigmoidinė	1024	32	32	32	32	32	32	32	32	480	32	1024
		Fiksuotas SLIC	sigmoidinė	672	32	32	32	32	32	32	32	32	32		

XGBoost

Registruoti	Daugiamačiai požymiai	Segmentavimo metodas	colsample bytree	gamma	max depth	min child weight	reg alpha
Neregistruoti	Visi požymiai	Proporcingas SLIC	0.412	0.005	6	0	0.530
		Fiksuotas SLIC	0.458	0.002	15	0	0.636
	Tik diskretūs taškai	Proporcingas SLIC	0.420	0.000	8	0	0.596
		Fiksuotas SLIC	0.457	0.247	13	0	0.008
Registruoti	Visi požymiai	Proporcingas SLIC	0.727	0.019	16	0	0.825
		Fiksuotas SLIC	0.665	0.080	13	1	0.446
	Tik diskretūs taškai	Proporcingas SLIC	0.586	0.020	6	0	0.046
		Fiksuotas SLIC	0.698	0.007	10	0	0.505

# Validacijos duomenų tikslumo matų medianos agreguotos konfigūracijos lygmenyje

		Registruoti	Daugiamačiai požymiai	Segmentavimo metodas	Preciziškumo vidurkis	Preciziškumo STD	Atitikimo vidurkis	Atitikimo STD	F1 vidurkis	F1 STD	Balansuoto tikslumo vidurkis	Balansuoto tikslumo STD	Specifiškumo vidurkis	Specifiškumo STD
Pilnai sujungtų sluoksnių neuroninis tinklas	Neregistruoti	Visi požymiai	Proporcingas SLIC	0.369	0.061	0.650	0.071	0.442	0.046	0.772	0.031	0.923	0.036	
			Fiksuotas SLIC	0.279	0.052	0.733	0.084	0.391	0.040	<b>0.831</b>	0.037	0.943	0.028	
		Tik diskretūs taškai	Proporcingas SLIC	0.353	0.074	0.633	0.091	0.413	0.059	0.784	0.044	0.936	0.037	
			Fiksuotas SLIC	0.332	0.041	0.738	0.114	0.395	0.043	0.812	0.056	0.942	0.016	
	Registruoti	Visi požymiai	Proporcingas SLIC	0.332	0.087	0.642	0.145	0.393	0.077	0.775	0.060	0.928	0.030	
			Fiksuotas SLIC	0.181	0.061	0.660	0.141	0.270	0.044	0.772	0.067	0.935	0.036	
		Tik diskretūs taškai	Proporcingas SLIC	0.282	0.078	0.686	0.113	0.364	0.077	0.782	0.052	0.911	0.036	
			Fiksuotas SLIC	0.186	0.047	0.667	0.076	0.268	0.038	0.801	0.021	0.933	0.017	
		Registruoti	Daugiamačiai požymiai	Segmentavimo metodas	Preciziškumas	Atitikimas	F1		Balansuotas tikslumas		Specifiškumas			
XGBoost	Neregistruoti	Visi požymiai	Proporcingas SLIC	0.40	0.750	0.471	<b>0.855</b>	0.959						
			Fiksuotas SLIC	0.50	0.600	0.540	0.794	0.985						
		Tik diskretūs taškai	Proporcingas SLIC	0.455	0.667	0.462	0.810	0.964						
			Fiksuotas SLIC	0.50	0.640	0.514	0.793	0.971						
	Registruoti	Visi požymiai	Proporcingas SLIC	0.571	0.650	0.533	0.780	0.971						
			Fiksuotas SLIC	0.45	0.560	0.483	0.748	0.976						
		Tik diskretūs taškai	Proporcingas SLIC	0.50	0.625	0.429	0.752	0.977						
			Fiksuotas SLIC	0.50	0.667	0.465	0.750	0.980						



# Validacijos duomenų tikslumo matų medianos agreguotos vienos konfigūracijos lygmenyje

Konfigūracija	Preciziškumas	Atitikimas	F1	Balansuotas tikslumas	Specifiškumas
Neregistruoti	0.382	0.675	0.446	<b>0.804</b>	0.951
Registruoti	0.361	0.657	0.431	<b>0.773</b>	0.948
Visi požymiai	0.389	0.667	0.457	<b>0.780</b>	0.948
Tik diskretūs taškai	0.356	0.667	0.426	<b>0.789</b>	0.951
Proporcingas SLIC	0.394	0.667	0.450	<b>0.783</b>	0.948
Fiksuotas SLIC	0.355	0.667	0.431	<b>0.792</b>	0.953
XGBoost	0.500	0.652	0.502	<b>0.792</b>	0.973
Pilnai sujungtų sluoksnių neuroninis tinklas	0.303	0.675	0.366	<b>0.785</b>	0.934



# P-reikšmės tarp lyginamų konfigūracijų

- Pilnai sujungtų sluoksnių neuroninis tinklas ir XGBoost modelis - 0.289765.
- Registruoti funkciniai duomenys su neregistruotais - 0.000007.
- Visi ištraukti požymiai iš funkcinių duomenų ir tik diskretūs taškai - 0.203341.
- Fiksuotas SLIC region skaičius ir proporcingas - 0.502053.

# Conclusions

- XGBoost klasifikavimo algoritmas pasiekia šiek tiek geresnius rezultatus nei pilnai sujungtų sluoksnių neuroninis tinklas. Tačiau, palyginimas yra statistiškai nereikšmingas.
- Neregistruoti funkciniai duomenys pasiekia statistiškai reikšmingai tikslesnius rezultatus nei registruoti.
- Nėra statistiškai reikšmingo skirtumo tarp visų požymių ir tik diskrečių taškų naudojimo apmokymui.
- Proporcingas skaičius SLIC regionų pasiekia geresnius rezultatus nei fiksuotas kai yra naudojamas XGBoost klasifikavimo algoritmas. Tačiau šis palyginimas yra statistiškai nereikšmingas.



# Kito pusmečio darbo planas

- Išlaikyti dalykų egzaminus:
  - Fundamentalieji informatikos ir informatikos inžinerijos metodai
  - Statistinis modeliavimas ir stochastinis optimizavimas
- Sudalyvauti bent vienoje tarptautinėje arba nacionalinėje konferencijoje pristatant disertacijos teorinio tyrimo rezultatus.