



**Vilnius
universitetas**

**Vilnius
universitetas**

**Algoritmų, leidžiančių
nustatyti medžiagos
cheminę sudėtį, iš hiper-
spektrinių duomenų tyrimas**

Vytautas Paura

Darbo vadovas: Dr. Virginijus Marcinkevičius

Darbo konsultantas: Dr. Valdas Rapševičius

Doktorantūros pradžios ir pabaigos metai: 2020–2024

Studijų planas

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2020/2021)	2	2	1	1	1	1 (konferencijoje)	
II (2021/2022)	2	2	1	1	1	1 (konferencijoje)	
III (2022/2023)			1	0	1	0	
IV (2023/2024)			1	0	1	0	

Pusmečio ataskaita

Egzaminai		
Planas	Įvykdyta	Būklė
Netiesiniai statistikos modeliai masinių duomenų analizėje	2022-06-23	Egzaminas <u>išlaikytas</u> .

Dalyvavimas konferencijose		
Planas	Įvykdyta	Konferencijos tipas
EEML 2022, Eastern European Machine Learning Summer School, Vilnius.	Pristatytas plakatas: V. Paura “Hyperspectral data synthesis using deep neural networks”	Tarptautinė <u>vasaros mokykla</u>

Disertacijos rengimo etapai

Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):

1. Ištirti klasikinius medžiagų cheminės sudėties nustatymo iš hiperspektrinių duomenų metodus.
2. Ištirti neneigiamų matricų faktorizavimo, mišinių modelių, sąsukos (angl. convolution) neuroninių tinklų, giliųjų neuroninių tinklų metodus hiperspektrinių duomenų analizei teorinį taikymą.
3. Ištirti šiuolaikinių mašininio mokymosi ir giliųjų neuroninių tinklų metodų panaudojamumą medžiagų cheminės sudėties nustatymui.

Tyrimo objektas

Tyrimo objektas:

- Hiper-spektriniai vaizdai ir medžiagų cheminės sudėties iš hiper-spektrinių vaizdų nustatymo algoritmai.

Sprendžiamos problemos:

- Grynujų pikselių skaičius nustatytas iš pateikto hiper-spektrinio vaizdo ar duomenų rinkinio.
- Randant grynujų medžiagų spektrines žymes (angl. signatures).
- Suskaičiuojant atrastų grynujų medžiagų sumaišymo proporcijas visame hiper-spektriniame vaizde.

2022/2023 m. m. darbo planas

- Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):
 - Problemų, kylančių iš tikslo, suformulavimas būsimiems eksperimentiniams ir analitiniams tyrimams.
 - Hiper-spektrinių duomenų rinkinių, skirtų cheminių medžiagų nustatymo algoritmams tirti, apžvalga ir prieinamumo tyrimas.
 - Hiper-spektrinių duomenų rinkinio, naudojamo cheminių medžiagų nustatymui, sudarymo metodikos kūrimas ir algoritmų optimizavimas.
 - Medžiagų cheminės sudėties nustatymo algoritmų eksperimentinis tyrimas panaudojant naujai sukurtą hiper-spektrinių duomenų rinkinį.
- Publikacijų rengimas:
 - Algoritmų palyginimo tyrimo rezultatų publikavimas (recenzuojamame leidinyje, CA WoS su Impact Factor).
 - Hiper-spektrinių duomenų rinkinio sudarymas, tyrimas ir publikacijos kūrimas iš šių duomenų.
 - Hiper-spektrinių duomenų analizės algoritmo kūrimas.

Cheminės sudėties nustatymas

Sudėties nustatymo algoritmų tipai:

1. Neneigiamų matricių faktorizavimas.
2. Autoenkoderių tinklai.
3. Retų duomenų regresija.

Metrikos:

1. Vidutinės kvadratinės paklaidos šaknis (Root mean squared error - RMSE)
2. S rekonstravimo paklaida (Signal reconstruction error SRE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}$$

$$SRE = 10 \log_{10} \left(\frac{E[\|x\|_2^2]}{E[\|x - \hat{x}\|_2^2]} \right)$$

Cheminės sudėties nustatymo eksperimentas

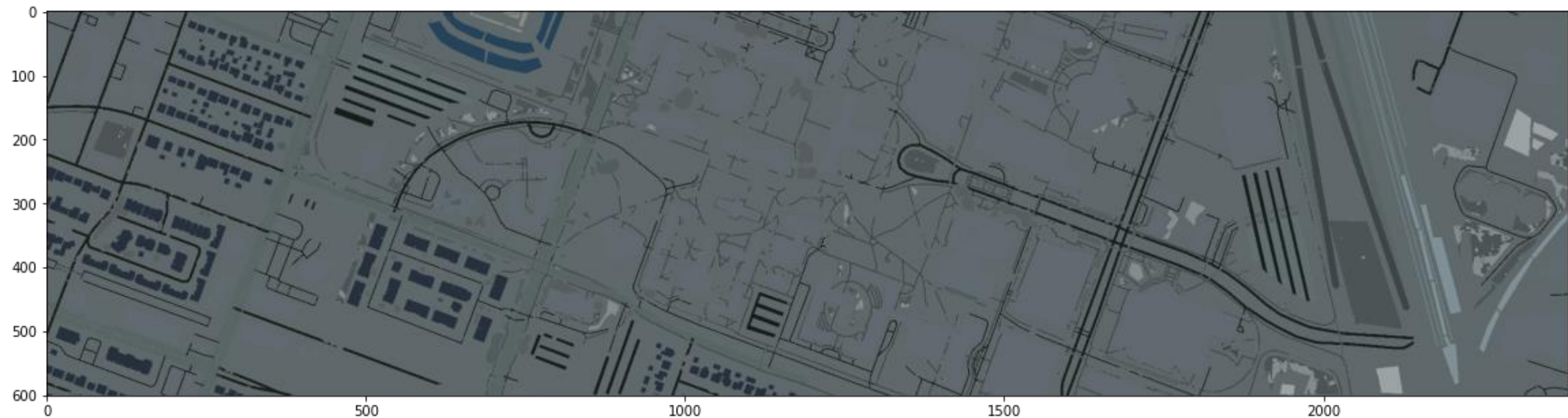
Duomenų rinkiniai:

- Dirbtiniai duomenų kubai sudaryti iš USGS spektrinės bibliotekos duomenų (Raymond F. Kokaly et al., 2017)
- “2018 IEEE GRSS data fusion” konkurso klasifikuotas miesto vaizdo paviršiaus elementų fragmentas.

Etaloninis eksperimentas:

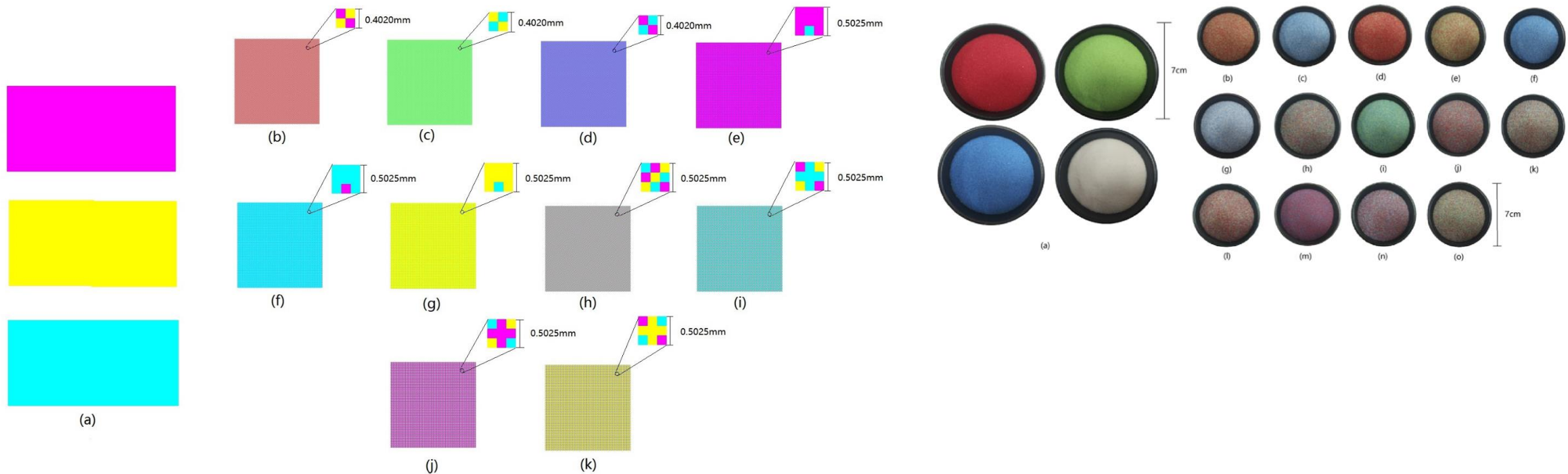
1. Algoritmų atsparumas medžiagų kiekio pasikeitimams.
2. Algoritmų atsparumas triukšmui.
3. Algoritmų gebėjimas atskirti medžiagas priklausomai nuo hiper-spektrinio vaizdo dydžio.

Cheminės sudėties nustatymo eksperimentas



“2018 IEEE GRSS data fusion” duomenų rinkinio fragmentas naudotas, kaip hiperspektrinių vaizdų generavimo pagrindas.

Esami duomenų rinkiniai

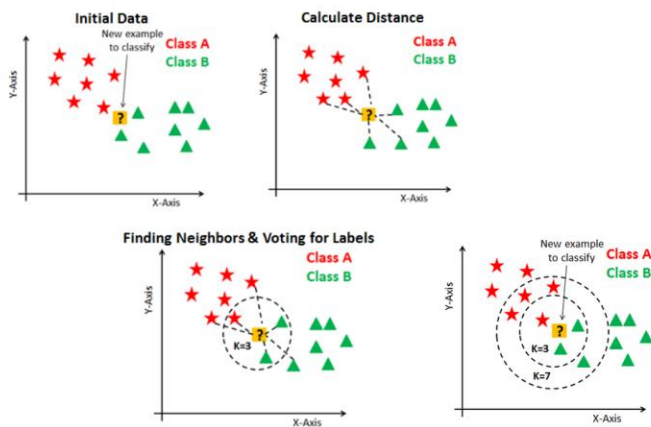


Šaltinis:

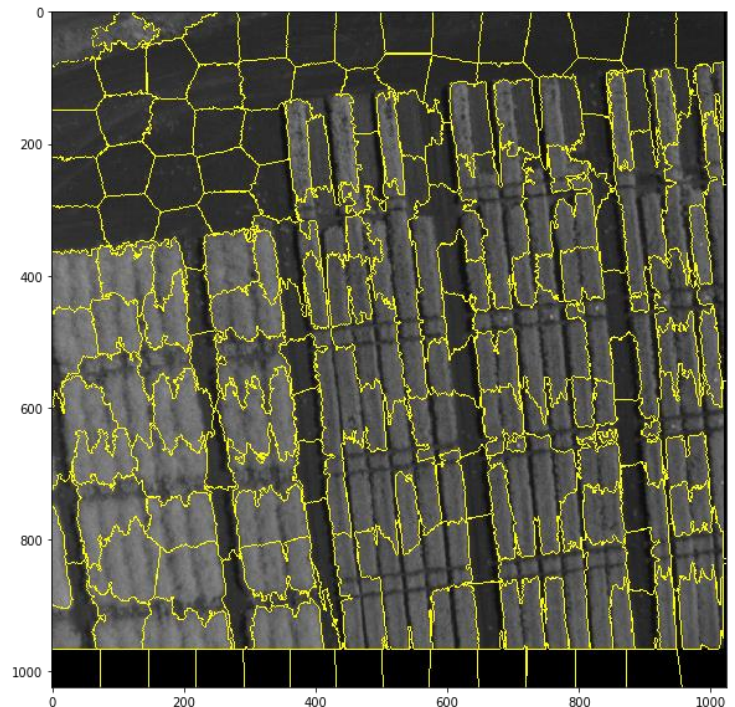
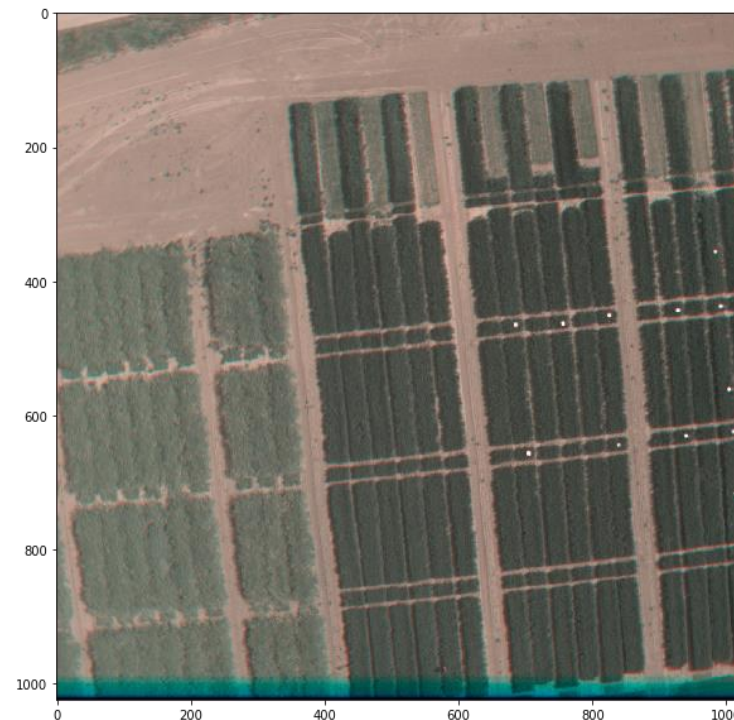
M. Zhao, J. Chen, and Z. He, "A laboratory-created dataset with ground-truth for hyperspectral unmixing evaluation," CoRR, vol. abs/1902.08347, 2019. [Online].

Duomenų analizė

Duomenų rinkinyje sudaromos klasės panaudojant KNN ir SLIC segmentavimo algoritmus pasirinktam ruožui.



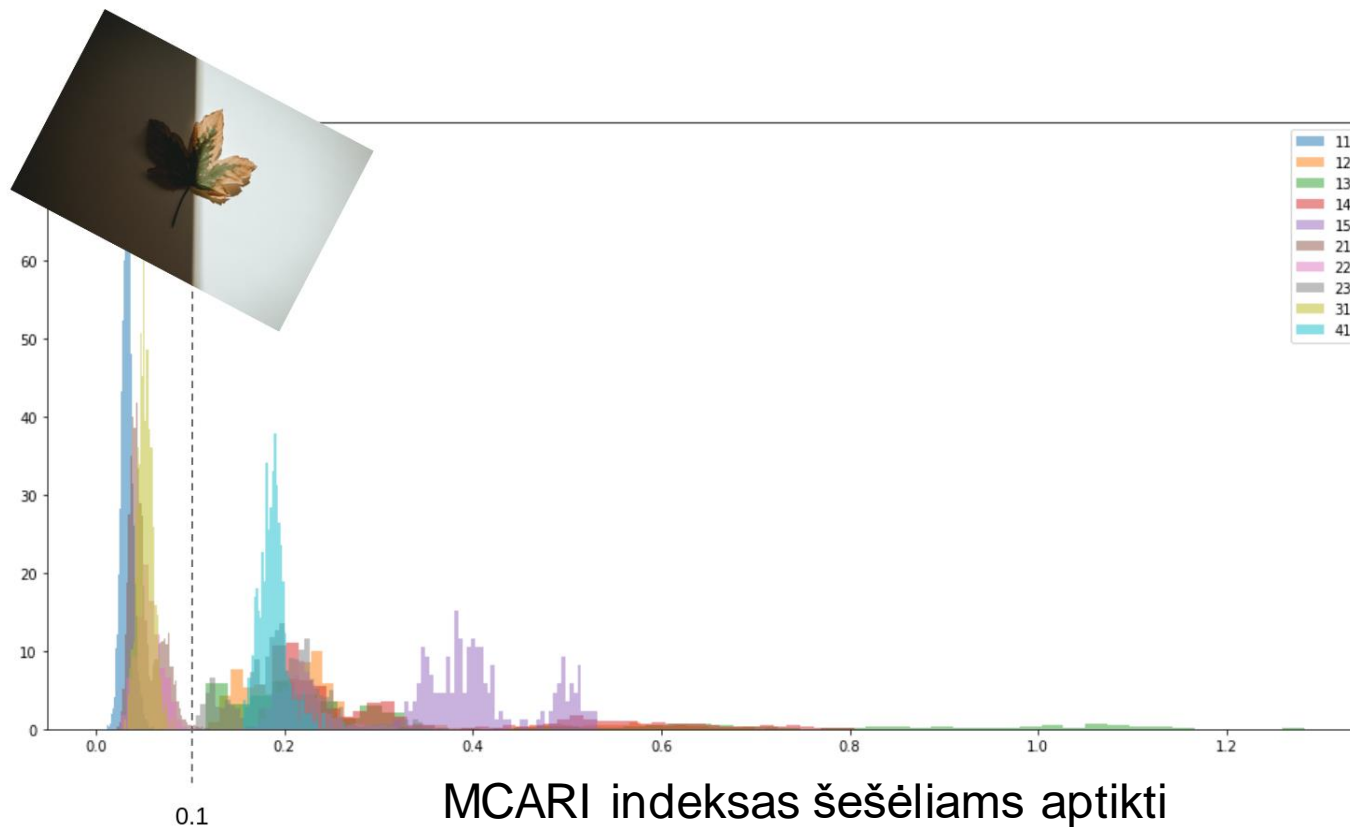
KNN



SLIC segmentavimas

Duomenų analizė

Pradinei analizei panaudojami vegetacijos indeksai:
NDVI, MCARI, CI



KNN ir SLIC nustatytos
klasės

Duomenų rinkinio sudarymo metodika

1. Pasirinktas hiperpsektinių duomenų kubas klasifikuojamas SLIC ir KNN algoritmais.
2. Panaudojant vegetacijos indeksus klasės patikslinamos, pašalinamos ekstremalios reikšmės.
3. Išgryninami etaloniniai piksleiai iš kiekvienos klasės naudojant nutriukšminimo metodus (SavGol) ir skaičiuojant vidurkius.

Išvados

- Yra didelis atvirai prieinamų hiper-spektrinių duomenų trūkumas, ypač medžiagų nustatymui skirtų duomenų rinkinių.
- Straipsniuose naudojami keli tie patys arba sintetiniai duomenų rinkiniai.
- Hiper-spektriniai duomenys ir jautrūs apšvietimui ir kitiems triukšmams.
- Hiper-spektrinių duomenų apdorojimui reikalingi modifikuoti klasterizavimo algoritmai, kurie atsižvelgtų į daugiadimensinius duomenis.

Kito pusmečio planas

1. Hiper-spektrinių duomenų rinkinio kūrimas ir publikacijos kūrimas iš šių duomenų.
2. Hiper-spektrinių duomenų analizės algoritmo kūrimas.
3. Dalyvavimas tarptautinėje konferencijoje pristatant cheminių medžiagų algoritmų analizės etaloninį eksperimentą.
4. Sukurto duomenų rinkinio ir jo analizės publikavimas (recenzuojamame leidinyje, CA WoS su Impact Factor).



**Vilniaus
universitetas**

Kontaktai

Akademijos g. 4

LT-08663 Vilnius

+370 6 256 79560

vytautas.paura@mif.stud.vu.lt