

Dirbtinio intelekto metodų tyrimas astrofizikinių objektų klasifikavimui ir/arba svarbių savybių nustatymui

Doktorantas: Tomas Mūžas

Darbo vadovas: prof. dr. Tadas Meškauskas

Darbo konsultantas: dr. Andrius Vytautas Misiukas Misiūnas

Doktorantūros studijų laikotarpis: 2022 – 2026 m.

Studijų planas ir jo vykdymo suvestinė

| Metai | Egzaminai | | Dalyvavimas konferencijose | | | | Publikacijos | | |
|-----------------|-----------|----------|----------------------------|----------|--------------|----------|-------------------|----------|------------------------|
| | | | Tarptautinės | | Nacionalinės | | Su citav. rodikl. | | |
| | Planas | Įvykdyta | Planas | Įvykdyta | Planas | Įvykdyta | Planas | Įvykdyta | Būsena |
| I (2022/2023) | 2 | 2 | - | 1 | - | - | - | - | - |
| II (2023/2024) | 2 | 1 | - | - | - | - | - | - | Baigiama rašyti |
| III (2024/2025) | - | - | 1 | - | - | - | 1 | - | - |
| IV (2025/2026) | - | - | 1 | - | - | - | 1 | - | - |
| Iš viso | 4 | 3 | 2 | 1 | 0 | 0 | 2 | 0 | - |

2022/2023 m.m. pavasario semestro darbo planas ir jo vykdymas (1)

Egzaminai

| Planas | Įvykdyta | Būklė |
|---|------------|-----------|
| Vaizdų ir duomenų analizė (FF) (2023-06-14) | 2023-06-14 | Išlaikyta |
| Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika (2023-06-22) | 2023-06-22 | Išlaikyta |
| Fundamentalieji informatikos ir informatikos inžinerijos metodai (2024-01-24) | 2024-01-24 | Išlaikyta |

Tyrimo objektas

- Astrofizikinių objektų klasifikavimas ir/ar savybių nustatymas
- profesionalų parengti arba savanorių balsavimu grįsti duomenų rinkiniai
- antžeminių ir kosmoso teleskopų nuotraukos

Tyrimo tikslas

Pasiūlyti naują arba patobulintą astrofizikinių objektų klasifikavimo modelį, paremtą dirbtinio intelekto metodais

Tyrimo uždaviniai

1. Sukurti astrofizikinių objektų duomenų analizės metodiką, apibrėžti kriterijus objektų klasifikavimo ir/ar savybių patikimumui įvertinti. Taip pat parengti metodiką sintetinių duomenų generavimui.
2. Atrinkti patikimų duomenų rinkinius, skirtus modelių apmokymui ir validavimui.
3. Apmokyti skirtingus dirbtinio intelekto modelius naudojant parengtą apmokymo duomenų rinkinį. Skaičiavimus atlikti naudojant paskirstyto skaičiavimo resursus bei grafinius procesorius.
4. Išskirti efektyviausius modelius bei jų parametrus, optimizuoti jų taikymą.

Pagrindiniai galaktikų tipai

Spiralinė



Eliptinė



Problemos

- Atlikus literatūros apžvalgą paaiškėjo, kad nors ir kiti autoriai pasiekia puikų tikslumą (95-99%), tačiau **naudojami tik nedideli duomenų rinkiniai (< 150,000)**.
- Cheng et al., apmokę modelį vos su 3,000 nuotraukų, bando klasifikuoti 20 mln. nuotraukų.
- Todėl kyla klausimas, kokio dydžio apmokymo duomenų imtis yra pakankama tiksliai klasifikuoti didelius duomenų kiekius

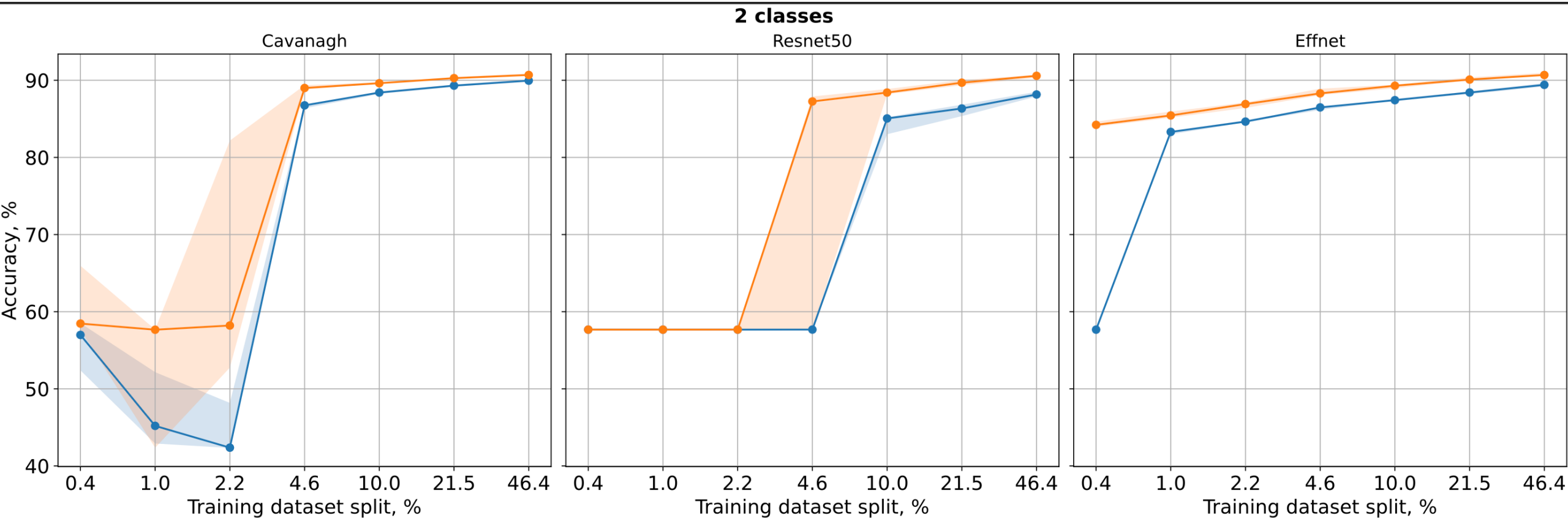
Duomenų rinkiniai

- Buvo paruošti trys dideli duomenų rinkiniai – 2 (820k galaktikų), 3 (781k) ir 5 (737k) klasių.
- Duomenų rinkiniai gauti apjungus visus savanorių balsavimų grįstus Galaxy Zoo (GZ) projektų duomenis
- Duomenų rinkiniai buvo filtruojami siekiant užtikrinti unikalias ir patikimas klasifikacijas:
 - Galaktikos apjungtos pagal coordinates
 - Galaktikos klasė visuose GZ projektuose turi sutapti

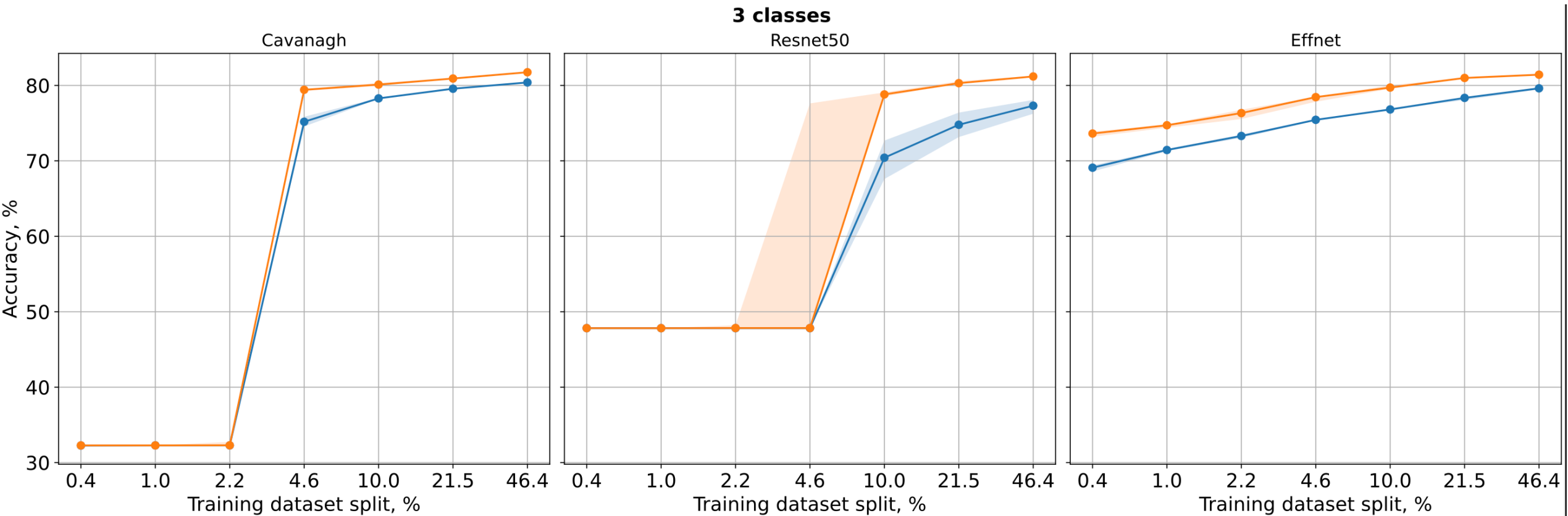
Atlikti eksperimentai

- Apmokymui naudojama tik labai nedidelis kiekis duomenų, visi likę – testavimui
- Apmokymo imtys: 0.4%, 1%, 2.2%, 4.6%, 10%, 21.5%, 46.2% nuo atitinkamo duomenų rinkinio dydžio
- Apmokyti trijų skirtingo sudėtingumo architektūrų modeliai
- Buvo vertinama ir augmentacijų įtaka

Esminiai rezultatai (2 klasės)



Esminiai rezultatai (3 klasės)



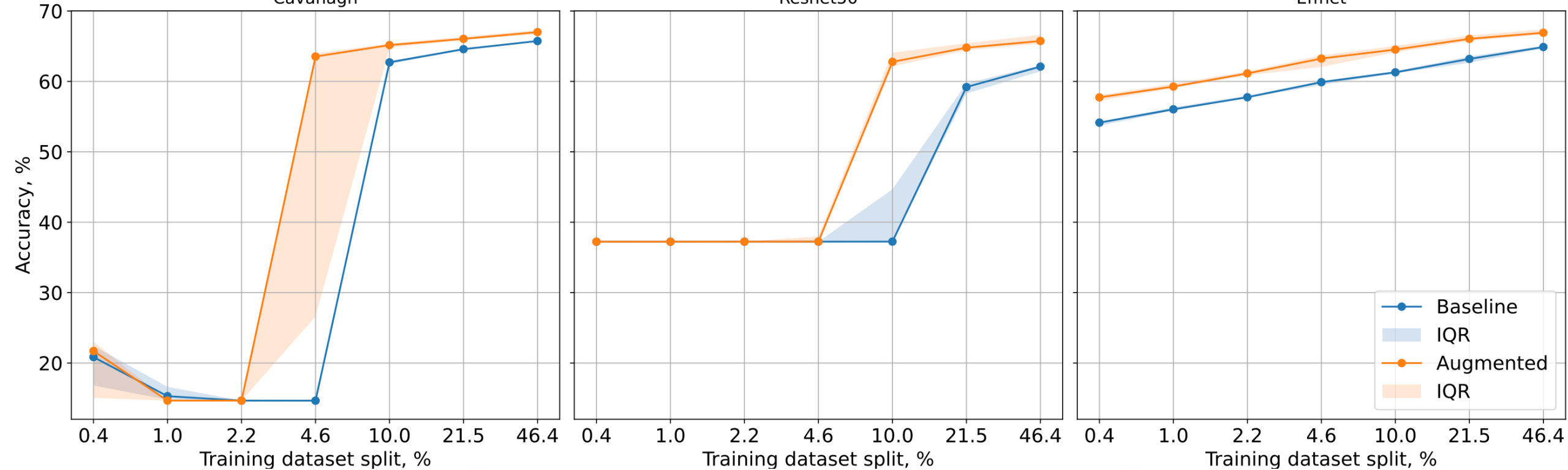
Esminiai rezultatai (5 klasės)

5 classes

Resnet50

Effnet

Cavanagh



Išvados

- Mūsų atliktuose eksperimentuose, siekiant gauti modelį, gebantį stabiliai klasifikuoti didelį kiekį galaktikų, reikia bent 72,704 nuotraukų 5 klasėms, 76,800 nuotraukų 3 klasėms, ir 80,896 nuotraukų 2 klasėms, kas atitinka 10% duomenų rinkinio dydžio.
- Visiems modeliams, standartinių augmentacijų technikų taikymas prilygsta apmokymo imties padidinimui 2 ar net daugiau kartų. Ši tendencija stebima visuose – 2, 3 ir 5 klasių – duomenų rinkiniuose.
- Iš visų tirtų modelių, EfficientNetV2S parodė stabiliausius rezultatus – persimokymas įvyko tik naudojant 2 klasių duomenų rinkinį ir mažiausią, 4,096 nuotraukų imtį.

Numatomi darbai

- Baigti rengti straipsnį ir jį publikuoti citavimo rodiklį turinčiame žurnale
- Pagerinti klasifikavimo rezultatus sukuriant naujas modelių architektūras ir/ar augmentacijos būdus

Ačiū už dėmesį!