



Kauno technologijos universitetas

Informatikos fakultetas

**Multi-klasijų mados vaizdų atpažinimas naudojant gilųjį
mokymąsi ir GAN grįstą duomenų sintezę**

Baigiamasis magistro projektas

Agnė Moleikaitytė

Projekto autorė

Prof. dr. Agnė Paulauskaitė-Tarasevičienė

Vadovė

Kaunas, 2025



Kauno technologijos universitetas

Informatikos fakultetas

**Multi-klasių mados vaizdų atpažinimas naudojant gilųjį
mokymąsi ir GAN grįstą duomenų sintezę**

Baigiamasis magistro projektas

Dirbtinio intelekto informatika (6211BX007)

Agnė Moleikaitytė

Projekto autorė

**Prof. dr. Agnė Paulauskaitė-
Tarasevičienė**

Vadovė

Prof. dr. Alfonsas Misevičius

Recenzentas

Kaunas, 2025



Kauno technologijos universitetas

Informatikos Fakultetas

Agnė Moleikaitytė

Multi-klasių mados vaizdų atpažinimas naudojant gilųjį mokymąsi ir GAN grįstą duomenų sintezę

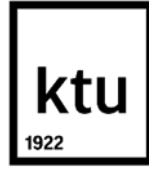
Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Agnė Moleikaitytė

Patvirtinta elektroniniu būdu



Kauno technologijos universitetas

Informatikos fakultetas

Baigiamojo magistro projekto užduotis

Projekto tema

Multi-klasių mados vaizdų atpažinimas naudojant gilųjį mokymąsi ir GAN grįstą duomenų sintezę

Reikalavimai ir sąlygos
(tikslinti pavadinimą
pagal poreikį)

Vadovas / Vadovė

(vadovo pareigos, vardas, pavardė, parašas)

(data)

Moleikaitytė Agnė. Multi-klasių mados vaizdų atpažinimas naudojant gilųjų mokymąsi ir GAN grįstą duomenų sintezę. Magistro baigiamasis projektas vadovė prof. dr. Agnė Paulauskaitė-Tarasevičienė; Kauno technologijos universitetas, informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Informatikos mokslai, Informatika (B01).

Reikšminiai žodžiai: GAN, StyleGAN, gilusis mokymasis, neuroniniai tinklai, klasifikavimas.

Kaunas, 2025. 60 p.

Santrauka

Šiame darbe nagrinėjama drabužių klasifikacijos problema nesubalansuotame DeepFashion duomenų rinkinyje. Apžvelgiami esami drabužių klasifikavimo metodai, pateikiamas patobulintas metodas, apjungiantis mažųjų klasių išplėtimą panaudojant generatyvinius priešiškaisius neuroninius tinklus (GAN) ir YOLO11 klasifikavimo modelį.

Pagerinti rezultatai lyginami su kitais balansavimo metodais ir modelių variacijomis. Pradinis 289,222 duomenų rinkinys, turintis 46 klases, buvo apdorotas – mažiausios klasės apjungtos į kitas arba pašalintos, rezultate paliekant 24 klases. Visi vaizdai apkirpti pagal rūbo ribinį langelį ir jų dydis pakeistas į 256x256px raišką. Tirti du balansavimo būdai: standartinės augmentacijos ir sintetinių duomenų generavimas naudojant StyleGAN2-ADA ir naujesnę StyleGAN3 modelio versiją. StyleGAN3 apmokymas konvergavo greičiau ir paliko mažiau artefaktų paveikslėliuose, nei StyleGAN2-ADA modelis.

Klasifikavimo dalyje vertintas linijinis YOLO11 modelis ir jo hierarchinės versijos. Originaliame rinkinyje bazinis modelis pasiekė 76,15% top-1, 92,59% top-3 ir 96,60% top-5 tikslumą. Išplėtus mažąsias klases StyleGAN3 sugeneruotais vaizdais, rezultatai pakilo iki 77,53% top-1, 93,48% top-3 ir 97,40% top-5. Šis GAN išplėsto duomenų rinkinio metodas pagerino augmentacijomis išplėstus ir hierarchinius YOLO11 metodus ir pagerino ankstesnius literatūroje apžvelgtus DeepFashion rezultatus (91,99% top-3, 96,44% top-5).

Rezultatai rodo, kad klasių pasiskirstymas turi didelės įtakos tikslumui. Išplėtus mažųjų klasių duomenis aukštos kokybės GAN paveikslėliais, pasiekiami geriausi rezultatai, išlaikant paprastą tinklo architektūrą. Projektavimo skyriuje aprašyti duomenų paruošimo ir architektūros sprendimai, eksperimentų skyriuje pateikti gauti eksperimentiniai rezultatai.

Moleikaitytė Agnė. Multi Class Fashion Detection Using Deep Learning and GAN Based Data Synthesis. Master's Final Degree Project; Prof. PhD. Agnė Paulauskaitė-Tarasevičienė; Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Computer science, Informatics (B01).

Keywords: GAN, StyleGAN, deep learning, neural networks, classification.

Kaunas, 2025. 60.

Summary

This study addresses the problem of clothing classification in the unbalanced DeepFashion dataset. It reviews existing methods for clothing classification and presents an improved pipeline that combines minority classes expansion with generative adversarial networks (GANs) and a YOLO11 classifier.

The improvement is highlighted by a comparison of balancing strategies and other model variants. For the experiments, the original dataset of 289 222 garment images assigned with 46 labels was cleaned. The smallest classes were merged or discarded, leaving 24 classes. Every image was cropped to its bounding box and resized to 256x256 px. Two balancing routes were examined: standard augmentations and synthetic data generation using StyleGAN2-ADA and the newer version – StyleGAN3. StyleGAN3 converged faster and produced fewer artifacts than StyleGAN2-ADA.

The classification stage evaluated a linear YOLO11 model alongside hierarchical variants. On the raw, imbalanced data the linear baseline achieved 76.15% top-1, 92.59% top-3 and 96.60% top-5 accuracy. Expanding minority classes with StyleGAN3 generated images raised performance to 77.53% top-1, 93.48% top-3 and 97.40% top-5. This GAN-balanced method outperformed both the augmentation-only and hierarchical YOLO11 methods, and surpassed the previously reported results on DeepFashion (91.99% top-3, 96.44% top-5).

Results indicate that class distribution highly affects accuracy. Expanding smaller classes data with high quality GAN images delivers the best results while keeping the network simple. The project design section details the data preparation and modelling choices, while the experiments section presents the experiment results.

Turinys

Lentelių sąrašas	9
Paveikslų sąrašas	10
Santrumpų ir terminų sąrašas	11
Įvadas.....	12
1. Drabužių vaizdų klasifikavimo užduoties analizė	13
1.1. Esamų sprendimų analizė	13
1.1.1. <i>FashionNet</i>	13
1.1.2. <i>BCRNN</i> modelio pritaikymas	13
1.1.3. <i>TBFS-Net</i>	14
1.1.4. Tekstūra ir forma grįstas dviejų srautų tinklas	14
1.1.5. Orientaciniais taškais grįstas konvoliucinis neuroninis tinklas	15
1.1.6. Pusiau prižiūrimasis dalijimosi žiniomis metodas	16
1.1.7. <i>AlexNet</i> pritaikymas.....	16
1.1.8. Atsitiktinių miškų pritaikymas	17
1.2. Kompiuterinės regos metodų analizė	17
1.2.1. CNN metodai.....	17
1.3. <i>YOLO</i>	18
1.3.1. <i>YOLO11</i>	19
1.4. GAN metodai.....	20
1.4.1. <i>StyleGAN</i>	21
1.4.2. <i>StyleGAN2</i>	23
1.4.3. <i>StyleGAN2-ADA</i>	23
1.4.4. <i>StyleGAN3</i>	23
2. Sprendimo projektavimas mados vaizdų atpažinimui.....	25
2.1. Reikalavimų specifikacija	25
2.1.1. Sistemos paskirtis	25
2.1.2. Tyrimo planas	25
2.1.3. Projekto kūrimo įrankiai ir aplinka.....	26
2.1.4. Funkciniai reikalavimai	26
2.1.5. Vertinimo metrikos.....	27
2.1.6. Nefunkciniai reikalavimai	28
2.2. Duomenų rinkinys	28
2.2.1. Duomenų rinkinio paruošimas	31
2.2.2. GAN modelių apmokymas duomenų sintezei.....	34
2.3. Hierarchinių modelių architektūra.....	37
2.3.1. Trijų lygių hierarchinis modelis	37
2.3.2. Dviejų lygių hierarchinis modelis	37
2.3.3. Trijų nepilnų lygių hierarchinis modelis	38
3. Eksperimentai atlikti mados vaizdų atpažinimo sprendimui.....	39
3.1. Linijinio <i>YOLOv11-cls</i> modelio eksperimentai su nepraplėstu duomenų rinkiniu	40
3.1.1. Linijinio modelio su pilnu duomenų rinkiniu rezultatai:.....	40
3.1.2. Linijinio modelio su apjungtu duomenų rinkiniu rezultatai:.....	41
3.2. Hierarchinio <i>YOLOv11-cls</i> modelio eksperimentai.....	42
3.2.1. Trijų lygių hierarchinio modelio rezultatai.....	43

3.2.2. Dviejų lygių hierarchinio modelio rezultatai.....	44
3.2.3. Trijų nepilnų lygių hierarchinio modelio rezultatai.....	46
3.3. Eksperimentai su GAN metodais išplestu apmokymo duomenų rinkiniu.....	47
3.3.1. <i>YOLOv11-cls</i> modelio eksperimentai su <i>StyleGAN2-ADA</i> metodo papildytu duomenų rinkiniui	47
3.3.2. <i>YOLOv11-cls</i> modelio eksperimentai su <i>StyleGAN3</i> metodo papildytu duomenų rinkiniui.	49
3.4. <i>YOLOv11-cls</i> modelio eksperimentai su augmentacijos metodais papildytu duomenų rinkiniui	50
3.5. Bendri rezultatai	52
Išvados	55
Literatūros sąrašas	56
Priedai.....	60
1 Priedas. Pristatytas mokslinis straipsnis 30-ojoje tarptautinėje informacinių technologijų konferencijoje „IVUS 2025“ Moleikaitytė A., Paulauskas J. (2025). Improving DeepFashion Dataset Classification Accuracy with StyleGAN2-ADA: Addressing Imbalanced Data in Fashion Image Recognition.....	60
2 Priedas. Dalyvavimo „IVUS 2025“ konferencijoje sertifikatas.	60
3 Priedas. M. Bacevicius, A. Paulauskaite-Taraseviciene, G. Zokaityte, L. Kersys, and A. Moleikaityte, “Comparative Analysis of Perturbation Techniques in LIME for Intrusion Detection Enhancement,” Mach Learn Knowl Extr, vol. 7, Mar. 2025, doi: 10.3390/make7010021.....	60

Lentelių sąrašas

1 lentelė. Duomenų rinkinio klasės, jų kiekis ir procentinė dalis	29
2 lentelė. Apjungtų klasių duomenų rinkinio klasės ir jų procentinė dalis	31
3 lentelė. Trijų hierarchinių lygių mazgų architektūra.....	37
4 lentelė. Dviejų hierarchinių lygių mazgų architektūra	38
5 lentelė. Trijų nepilnų hierarchinių lygių mazgų architektūra.....	38
6 lentelė. Linijinio modelio, klasifikuojančio neapjungtų klasių duomenų rinkinį, rezultatai	41
7 lentelė. Linijinio modelio, klasifikuojančio apjungtų klasių duomenų rinkinį, rezultatai	42
8 lentelė. Trijų lygių hierarchinio modelio mazgų rezultatai.....	43
9 lentelė. Trijų lygių hierarchinio modelio rezultatai.....	44
10 lentelė. Dviejų lygių hierarchinio modelio mazgų rezultatai	45
11 lentelė. Dviejų lygių hierarchinio modelio rezultatai.....	46
12 lentelė. Trijų nepilnų lygių hierarchinio modelio mazgų rezultatai.....	46
13 lentelė. Trijų nepilnų lygių hierarchinio modelio rezultatai.....	47
14 lentelė. Linijinio modelio su <i>StyleGAN2-ADA</i> generuotais duomenimis rezultatai	48
15 lentelė. Linijinio modelio su <i>StyleGAN3</i> generuotais duomenimis rezultatai.....	50
16 lentelė. Linijinio modelio su augmentuotais duomenimis rezultatai	50
17 lentelė. Linijinio modelio su augmentuotais duomenimis svoriniai tikslumo rezultatai	50
18 lentelė. Linijinio modelio su augmentuotais duomenimis makro tikslumo rezultatai	50
19 lentelė. Bendros modelių top-k metrikos	52
20 lentelė. Bendros modelių svorinės tikslumo metrikos	53
21 lentelė. Bendros modelių makro tikslumo metrikos	53

Paveikslų sąrašas

1 pav. <i>FashionNet</i> architektūra [1].....	13
2 pav. <i>BCRNN</i> modelio pritaikymo sprendimo architektūra [14].....	14
3 pav. <i>TBFS-Net</i> architektūra [16]	14
4 pav. Tekstūra ir forma grįstas dviejų srautų tinklo architektūra [20].....	15
5 pav. Orientaciniais taškais grįsto konvoliucinio neuroninio tinklo architektūra [21].....	16
6 pav. Pusiau prižiūrimojo dalijimosi žiniomis metodo architektūra [25].....	16
7 pav. CNN architektūra [30].....	18
8 pav. YOLO modelio versijų ir jų skirtingų dydžių mAP50-95 metrikų palyginimas [37].....	19
9 pav. <i>YOLO11</i> architektūra [39].....	20
10 pav. GAN architektūra	21
11 pav. <i>StyleGAN</i> generatoriaus architektūra [41].....	22
12 pav. Tyrimo schema	26
13 pav. <i>DeepFashion</i> duomenų rinkinio paveikslėlių pavyzdžiai [1].....	29
14 pav. Duomenų rinkinio klasių pasiskirstymas	31
15 pav. Apjungto duomenų rinkinio klasių pasiskirstymas	32
16 pav. Išplėsto duomenų rinkinio klasių pasiskirstymas.....	33
17 pav. <i>StyleGAN2-ADA</i> generatoriaus ir diskriminatoriaus praradimai apmokymo metu (<i>Blazer</i> klasė).....	34
18 pav. <i>StyleGAN2-ADA Blazer</i> klasės sugeneruotų paveikslėlių pavyzdys.....	35
19 pav. <i>StyleGAN3</i> generatoriaus ir diskriminatoriaus praradimai apmokymo metu (<i>Blazer</i> klasė).....	36
20 pav. <i>StyleGAN3 Blazer</i> klasės sugeneruotų paveikslėlių pavyzdys.....	36
21 pav. Apmokymo praradimai.....	39
22 pav. Validacijos praradimai	40
23 pav. Linijinio <i>YOLOv11</i> modelio, apmokyto su visomis duomenų rinkinio klasėmis, susimaišymo matrica	41
24 pav. Linijinio <i>YOLOv11</i> modelio, apmokyto su apjungtomis duomenų rinkinio klasėmis, susimaišymo matrica	42
25 pav. Trijų lygių hierarchinio <i>YOLOv11</i> modelio susimaišymo matrica	44
26 pav. Dviejų lygių hierarchinio <i>YOLOv11</i> modelio susimaišymo matrica	45
27 pav. Trijų nepilnų lygių hierarchinio <i>YOLOv11</i> modelio susimaišymo matrica	47
28 pav. Linijinio modelio su <i>StyleGAN2-ADA</i> generuotais duomenimis susimaišymo matrica	48
29 pav. Linijinio modelio su <i>StyleGAN3</i> generuotais duomenimis susimaišymo matrica	49
30 pav. Linijinio modelio su augmenteduotais duomenimis rezultatai	51

Santrumpų ir terminų sąrašas

Santrumpos:

CNN (angl. *Convolutional Neural Network*) – konvoliucinis neuroninis tinklas.

GAN (angl. *Generative Adversarial Network*) - generatyvinis priešiškas neuroninis tinklas.

MLP (angl. *Multi-Layer Perceptron*) – daugiasluoksnis perceptronas.

ADA - adaptyvi duomenų augmentacija.

SURF (angl. *Speeded-Up Robust Features*) - pagreitinotos patikimos savybės.

HOG (angl. *Histogram of Oriented Gradients*) - orientuotų gradientų histograma.

LBP (angl. *Local Binary Patterns*) - vietinės dvejetainės struktūros.

SSD (angl. *Sum of Squared Differences*) - kvadratinių skirtumų suma.

Įvadas

Drabužių automatizuotas apdorojimas tiesiogiai siejasi su pramonės skaitmenizavimu, todėl šiandien dienai tai yra aktualus uždavinys keliantis daug iššūkių. Šio uždavinio svarbą lemia plati sprendimo pritaikymo sritis: elektroninė ir mažmeninė prekyba drabužiais ir iš to kylantys iššūkiai: inventoriaus valdymas, automatinis drabužių informacijos sukėlimas į elektroninę erdvę, produktų rekomendacijų pateikimas naudotojui. Užduotį komplikuoja drabužiams būdingas stilių kintamumas, subtilūs skirtumai tarp kategorijų, įvairi drabužių nuotraukų kokybė, kuriose drabužiai gali būti atvaizduoti skirtingais kampais, ant žmogaus kūno, manekeno, pakabos ar kitaip.

Svarbus tyrimo aspektas – klasių disbalansas. Ši problema dažnai pasitaiko mados duomenų rinkiniuose, kuomet dominuojančios klasės turi tūkstančius paveikslėlių, tuo tarpu kitos klasės turi minimalų kiekį duomenų [1], [2]. Šis duomenų disbalansas lemia apmokyto modelio šališkumą – modelis apsimoko geriau prognozuoti dominuojančias klases, o mažas klases prognozuoja žemu tikslumu [3], [4], [5], [6]. Išsprendus šią problemą, drabužių duomenų skaitmenizavimo užduočių, tokių kaip inventoriaus valdymas, automatinis duomenų skaitmenizavimas, automatinis duomenų įkėlimas į elektronines komercines platformas, rekomendacijų personalizavimas, patikimumas pakiltų ir šios sistemos atneštų didesnę naudą.

Tradiciniai nesubalansuoto duomenų rinkinio problemos sprendimai, tokie kaip duomenų augmentacija, svorinės praradimų funkcijos pritaikymas dalinai išsprendžia šią problemą, tačiau šie sprendimai nepadedą iš tiesų įnešti pakankamai diversifikuotų duomenų mažoms klasėms. Šiai problemai spręsti, pasitelkiamas sintetinių duomenų generavimas, panaudojant generatyvinius priešiškuosius neuroninius tinklus (GAN) [7], [8], [9], [10].

Darbo tikslas – padidinti drabužių klasifikavimo tikslumą, pasitelkiant duomenų generavimo metodus klasių disbalanso problemai išspręsti.

Darbo uždaviniai:

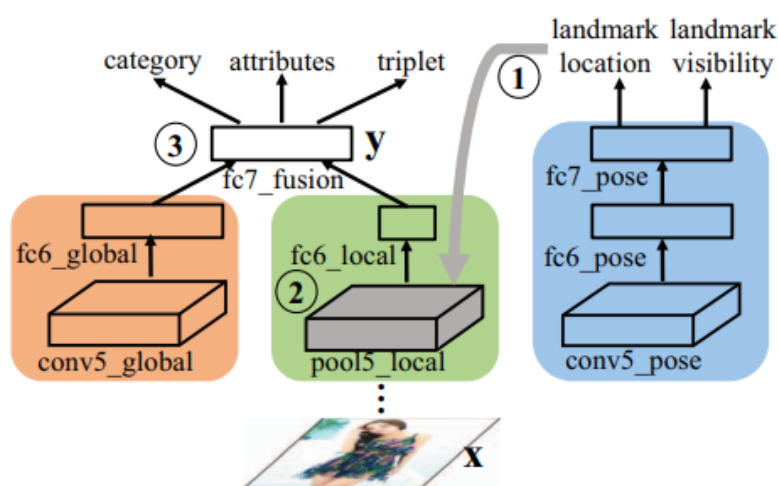
1. Atlikti esamų drabužių vaizdų klasifikavimo sprendimų literatūros analizę.
2. Atlikti duomenų generavimo metodų analizę.
3. Apdoroti ir paruošti eksperimentams turimą duomenų rinkinį.
4. Atlikti bandymus su giliojo mokymosi modeliais.
5. Atlikti eksperimentus ir apskaičiuoti modelio tikslumo metrikas.
6. Palyginti rezultatus ir pateikti išvadas.

1. Drabužių vaizdų klasifikavimo užduoties analizė

1.1. Esamų sprendimų analizė

1.1.1. FashionNet

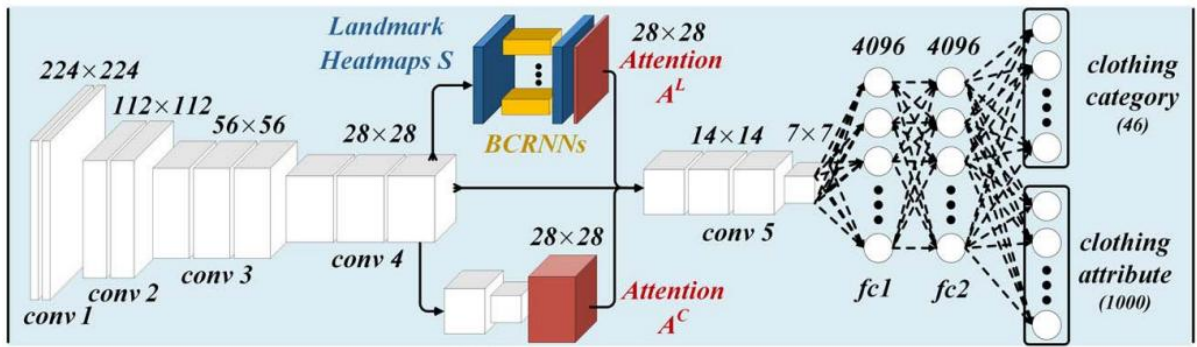
Šis sprendimas pristatytas 2016 metais, kartu su naujai sukurtu duomenų rinkiniu *DeepFashion* [1]. Sprendimo neuroninio tinklo architektūra yra panaši į *VGG-16* [11] – naudojami *VGG-16* konvoliucijos sluoksniai. Šie sluoksniai sujungiami su trijų srautų architektūra (žr. **1 pav.**), pritaikyta sprendžiamam uždaviniui. Srautuose naudojami skirtingi sluoksniai įvairioms drabužių savybėms užfiksuoti. Pirmajame sraute fiksuojamos globalios drabužių savybės, antrasis srautas atsakingas už savybių aptikimą, atsižvelgiant į drabužių orientacinius taškus, trečiasis srautas atsakingas už drabužių orientacinių taškų ir jų matomumo prognozavimą. Pirmojo ir antrojo srautų išvestys naudojamos drabužių kategorijoms ir atributams nuspėti. Siūlomo sprendimo drabužių klasifikavimo tikslumas matuojamas *top-3* ir *top-5* metrikomis. *Top-3* tikslumas siekia 82.58%, *top-5* – 90.17% [1].



1 pav. FashionNet architektūra [1]

1.1.2. BCRNN modelio pritaikymas

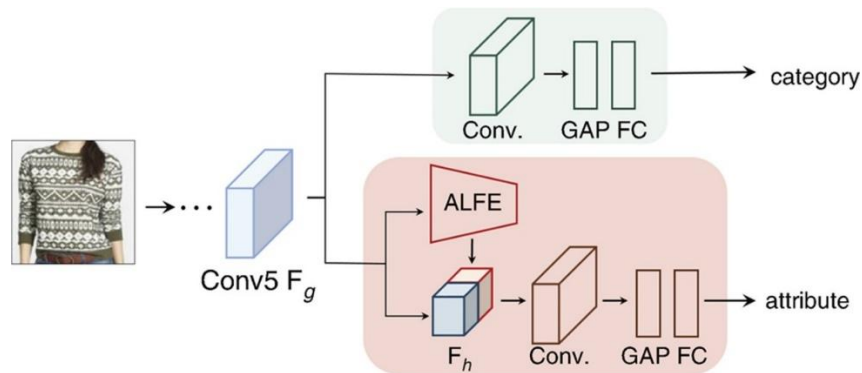
Sprendimas pasiūlytas 2018 metais. Sprendimas naudoja *DeepFashion* [1] duomenų rinkinį. Siūlomas dviejų krypčių konvoliucinio rekurentinio neuroninio tinklo (angl. *Bidirectional Convolutional Recurrent Neural Network*) architektūros panaudojimas. Šis sprendimas panaudoja *VGG-16* [11] konvoliucijos sluoksnius savybėms išgauti. Metodas išsišakoja į tris srautus (žr. **2 pav.**), kurių pirmasis išgauna orientacinių drabužių taškų šilumos žemėlapius (angl. *heatmaps*). Konvoliucijų išvestis ir šilumos žemėlapiai panaudojami kaip BCRNN įvestis orientacinių taškų savybėms išgauti. Antrasis srautas perduoda konvoliucijų išvestį tolesniam sluoksniui. Trečiajame sraute pritaikomas iš apačios į viršų iš viršaus į apačią tinklas (angl. *bottom-up top-down network*) [12][13]. Šiame sraute išgaunamas globalių drabužių savybių dėmesio žemėlapis (angl. *attention map*). Visų trijų srautų išvestys sujungiamos į vieną ir perduodamos tolesniam konvoliucijos sluoksniui ir vidurkių telkimo sluoksniams. Šiuo metodu pasiektas 90.99% *top-3* ir 95.78% *top-5* drabužių klasifikavimo tikslumas [14].



2 pav. BCRNN modelio pritaikymo sprendimo architektūra [14]

1.1.3. TBFS-Net

Sprendimas pristatytas 2019 metais. Sprendimas naudoja *DeepFashion* [1] duomenų rinkinį. Šiuo sprendimo būdu siekiama atsisakyti orientacinių taškų prognozavimo panaudojimo drabužiams klasifikuoti, kadangi ši užduotis didina modelio sudėtingumą ir skaičiavimų laiką. Siūloma architektūra – daugiafunkcinio mokymosi (angl. *multi-task learning*) modelis (žr. 3 pav.), kurį sudaro modelio pagrindinė dalis – VGG-16 [11] modelio konvoliucijų sluoksnis ir kita modelio dalis, kurią sudaro du srautai (angl. *streams*) kategorijoms ir atributams klasifikuoti. Abiejuose srautuose naudojami konvoliucijos sluoksniai, globalusis vidurkių telkimas (angl. *global average pooling*) ir visiškai sujungti sluoksniai. Į atributų srauto architektūrą įeina suaktyvintas vietinių savybių išgavėjas (angl. *activated local feature extractor*). Šis komponentas išgauna vietines savybes iš aktyvacijų žemėlapių, kuris gaunamas pasitelkiant *Grad-CAM* [15] metodą ir pasitelkiant globalųjį vidurkių telkimą, atrenka keletą regionų, kuriuose aktyvacijos reikšmės yra didžiausios. Šie regionai perduodami tolesniam modelio blokui, kuris nuspėja drabužių atributus. Siūlomo sprendimo drabužių klasifikavimo tikslumas matuojamas *top-3* ir *top-5* metrikomis. *Top-3* tikslumas siekia 91.37%, *top-5* – 95.26% [16].

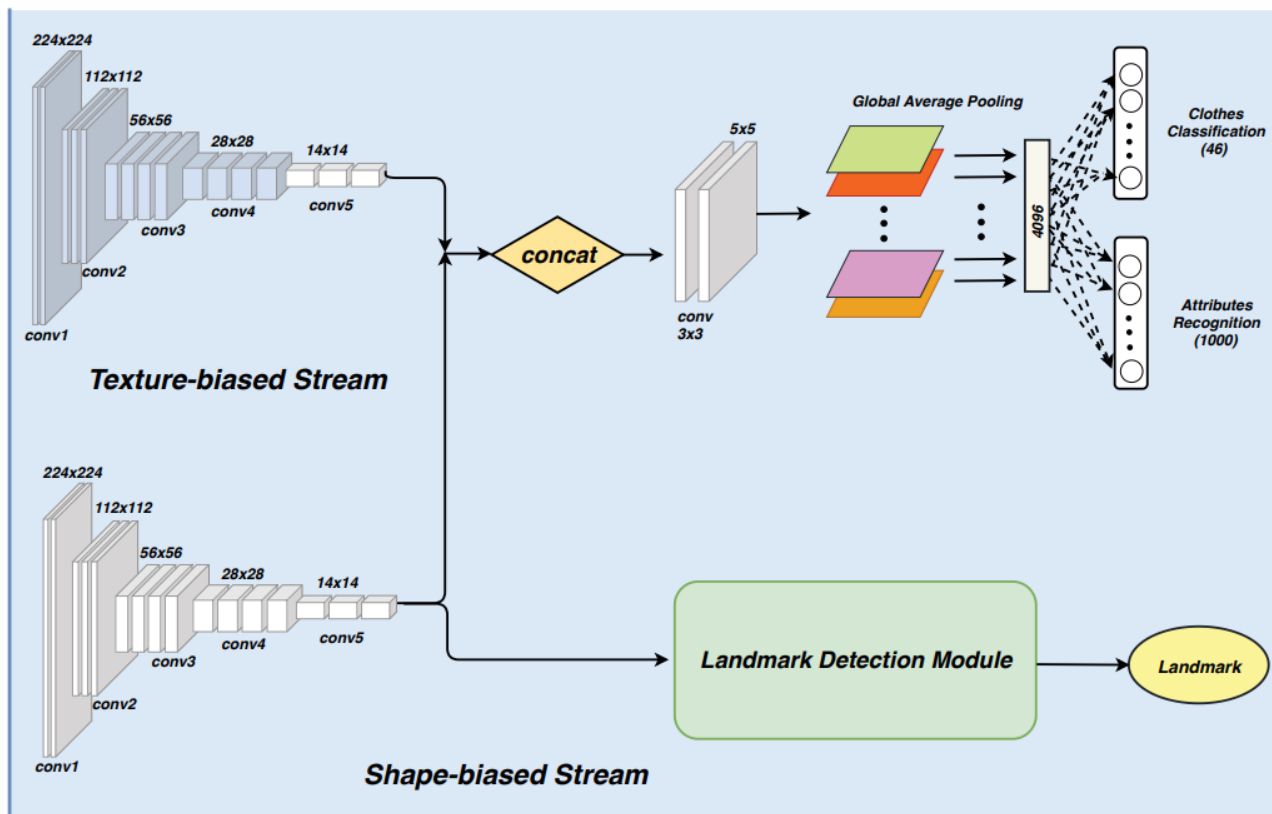


3 pav. TBFS-Net architektūra [16]

1.1.4. Tekstūra ir forma grįstas dviejų srautų tinklas

Šis sprendimas pasiūlytas 2020 metais. Drabužių klasifikavimui siūloma architektūra sudaryta iš dviejų srautų (žr. 4 pav.). Pirmasis srautas yra paremtas tekstūra, antrasis – forma. Tekstūros srautui naudojamas konvoliucinis neuroninis tinklas, apmokytas su *ImageNet* duomenų rinkiniu, kadangi klasikiniai konvoliuciniai neuroninių tinklų modeliai yra linkę klasifikuoti vaizdinę informaciją pagal tekstūrą ir formą [17][18][19]. Antrasis srautas yra skirtas drabužių orientacinių taškų, iš kurių galima spręsti apie drabužio formą aptikimui. Drabužių klasifikavimui ir atributų spėjimui panaudotas

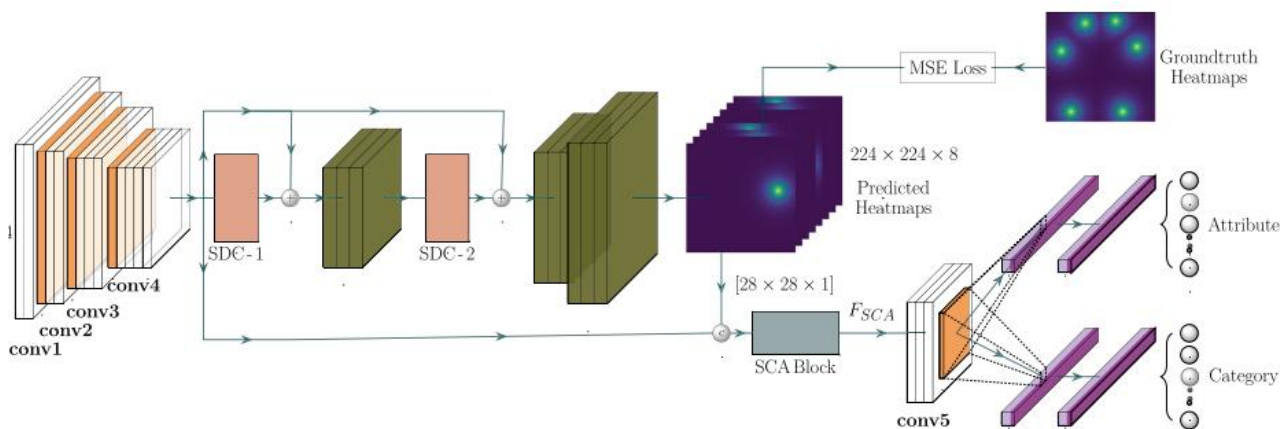
pirmasis srautas, kurio savybių žemėlapius sudarė sujungta *ImageNet* konvoliucinio neuroninio tinklo konvoliucijų išvestis kartu su antrojo srauto sluoksnių išvestimi. Orientaciniams taškams aptikti buvo panaudotas antrasis srautas (**4 pav.**). Šis sprendimo būdas išsprendžia konvoliucinių neuroninių tinklų šališkumą vaizdo tekstūrai [17][18][19]. Šiuo metodu pasiektas 91.99% *top-3* ir 96.44% *top-5* drabužių klasifikavimo tikslumas [20]. Sprendimas naudoja *DeepFashion* [1] duomenų rinkinį.



4 pav. Tekstūra ir forma grįstas dviejų srautų tinklo architektūra [20]

1.1.5. Orientaciniais taškais grįstas konvoliucinis neuroninis tinklas

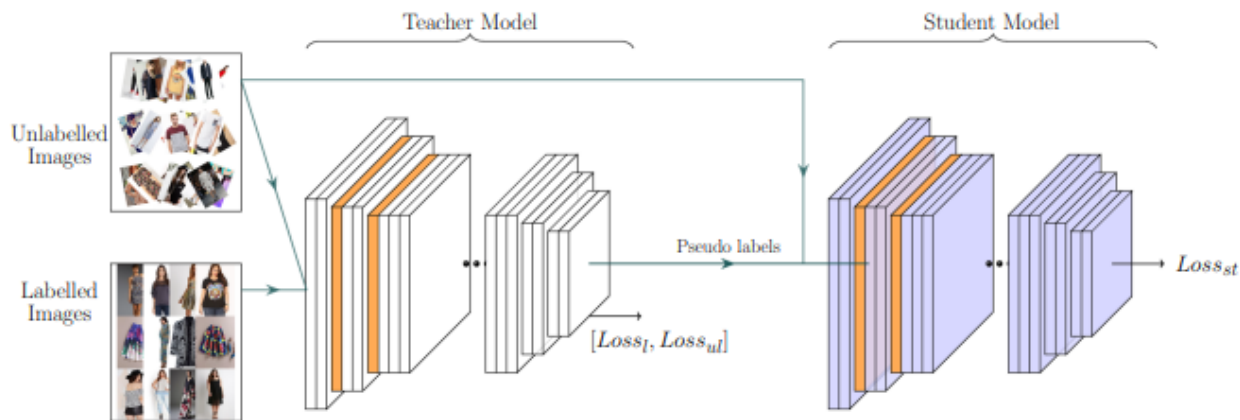
Sprendimas pristatytas 2020 metais. Sprendimas naudoja *DeepFashion* [1] duomenų rinkinį. Modelis naudoja *VGG-16* [11] konvoliucijos sluoksnius drabužių savybėms išgauti. Toliau srautas išsišakoja į dvi dalis (žr. **5 pav.**). Pirmasis srautas skirtas orientaciniams taškams išgauti, antrasis – atributams nuspėti ir drabužiams klasifikuoti. Pirmajame sraute išgaunami šilumos žemėlapiai, indikuojantys orientacinių taškų lokacijas, antrajame panaudojami pirmojo srauto šilumos žemėlapiai ir konvoliucijų sluoksnių išvestis, atributų ir kategorijai spėti. Šiuo metodu pasiektas 91.02% *top-3* ir 96.20% *top-5* drabužių klasifikavimo tikslumas [21].



5 pav. Orientaciniais taškais grįsto konvoliucinio neuroninio tinklo architektūra [21]

1.1.6. Pusiau prižiūrimasis dalijimosi žiniomis metodas

2022 metais pasiūlytas sprendimas, kuris pasitelkia pusiau prižiūrimąjį (angl. *semi-supervised*) mokymosi modelį (žr. 6 pav.). Modelį sudaro mokytojo ir mokinio pora. Modelis-mokytojas naudoja orientaciniais taškais grįsto konvoliucinio neuroninio tinklo architektūrą (žr. 1.1.5). Modelis-mokinys *ResNet-18* [22] architektūros modelis su papildomais dvejais visiškai sujungtais sluoksniais. Šiame metode naudojami keli duomenų rinkiniai: *DeepFashion* [1], *iMaterialist* [23], *Fashion10000* [24] duomenų rinkiniai. Dalis *Deepfashion* duomenų rinkinio panaudota modeliui-mokytojui, likusi dalis ir kiti duomenų rinkiniai – modeliui-mokiniui. Panaudojant šį metodą, pasiektas 91.06% *top-3* ir 96.35% *top-5* drabužių klasifikavimo tikslumas [25].



6 pav. Pusiau prižiūrimojo dalijimosi žiniomis metodo architektūra [25]

1.1.7. AlexNet pritaikymas

Šis sprendimas sukurtas perdirbamų drabužiams, slenkantiems konvejeriu klasifikuoti. Sistemą sudaro IoT įranga: „Raspberry pi“, kamera ir konvejeriu judantys drabužiai, kurie klasifikuojami į devynias kategorijas, panaudojant „AlexNet“ CNN modelį. Šis sprendimas skirtas pakeisti rankinį rūšiavimo darbą drabužių perdirbimo įmonėse. Šiam uždaviniui išspręsti, buvo naudojamas 2400 drabužių nuotraukų rinkinys, kuriame buvo pilnai matomi drabužiai ir 900 paveikslėlių, kuriuose kai kurios drabužių dalys nebuvo matomos. Šio sprendimo tikslumas – apie 74%, kai kamera užfiksuotas pilnas drabužis ir apie 68%, kai dalis drabužio nesimato[26].

1.1.8. Atsitiktinių miškų pritaikymas

Šis sprendimas pristatytas 2013 metais, pritaikytas viršutinės dalies drabužiams klasifikuoti, bei drabužių savybių išgavimui. Šiame sprendime, iš nuotraukos išgaunamos vaizdo savybės, tokios kaip SURF, HOG, LBP, SSD ir spalvų požymiai. Šios savybės apdorojamos ir perduodamos dviem klasifikatoriams. Pirmasis klasifikatorius prognozuoja drabužio klasę, antrasis drabužio atributus. Sprendime naudojama atsitiktinių miškų patobulinta architektūra. Sprendimo duomenų rinkinį sudaro 80 tūkst. nuotraukų. Drabužiai klasifikuojami į 15 klasių ir išskiriamos 8 atributų kategorijos, kurios iš viso turi 78 rūbų savybes. Sprendimo tikslumas – apie 41% [27].

1.2. Kompiuterinės regos metodų analizė

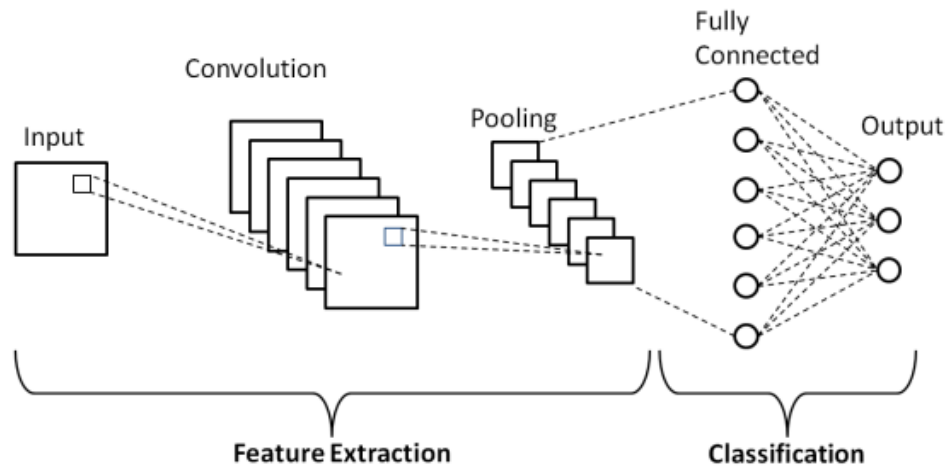
1.2.1. CNN metodai

Konvoliuciniai neuroniniai tinklai įprastai yra naudojami objektams aptikti, objektams klasifikuoti ir kitoms su kompiuterine rega susijusioms užduotims spręsti [28]. CNN yra įgyvendintas bandant atkartoti žmogaus smegenų veiklą, kurios metu išskiriamos vaizdo savybės, kurios yra svarbiausios tam tikrų užduočių sprendimui.

CNN architektūrą sudaro (žr. 7 pav.):

1. Įvesties sluoksnis, kuris priima neapdorotus vaizdo duomenis.
2. Konvoliuciniai sluoksniai, kurie su įvesties duomenimis atlieka matematinę operaciją, vadinamą konvoliucija, kurios metu iš vaizdo išgaunamos tam tikros savybės panaudojant filtrus. Šios operacijos metu apskaičiuojama filtro ir vaizdo dalies sandaugos reikšmės. Šis procesas kartojamas daugeliui skirtingų filtro padėčių, kol gaunamas naujas vaizdas, vadinamas savybių žemėlapiu.
3. Jungiamasis sluoksnis (angl. *pooling*), kuris naudojamas savybių žemėlapio matmenims sumažinti, kad būtų galima sumažinti reikalingas skaičiavimo sąnaudas. Mažinimas vykdomas pritaikant sluoksnių sujungimo operaciją.
4. Pilnai sujungti sluoksniai, kurie atsakingi už konvoliucinių ir jungiamųjų sluoksnių išgautų savybių interpretavimą. Pilnai sujungtus sluoksnius sudaro dirbtiniai neuronai, kurie naudoja iš vaizdo išskirtas savybes spėjimui pateikti.
5. Išvesties sluoksnis, kuris priima galutinį sprendimą apskaičiuodamas išvestį. Šis sluoksnis paprastai naudoja *Softmax* funkciją galutiniam rezultatui apskaičiuoti[29].

Įvairūs kiti dirbtinio intelekto metodai vaizdų analizei yra grįsti konvoliuciniais neuroniniais tinklais.



7 pav. CNN architektūra [30]

CNN modeliai gali būti pritaikomi drabužių segmentacijai, klasifikacijai, objektų aptikimui.

Segmentacija – tai metodas, kuris suskaido vaizdą į pogrupius, vadinamus vaizdo segmentais, kurie padeda sumažinti vaizdo sudėtingumą, supaprastindamas tolimesnį vaizdo apdorojimą ar analizę [31]. Visiems tai pačiai kategorijai priklausantiems vaizdo elementams arba pikseliams priskiriama bendra etiketė. Vėliau šiuos duomenis galima panaudoti norint supaprastinti įvestį, paduodama kitam vaizdo apdorojimo žingsniui, tik reikiamos vaizdo dalies išskyrimui, iš kurios norima išgauti informaciją.

Segmentacija naudinga tais atvejais, kai vaizduose yra daugiau nei vienas rūbas, pavyzdžiui: užfiksuotas žmogus, vilkintis daugiau nei vieną rūbą, o ne atskiras rūbas. Segmentacija yra tinkamas pasirinkimas norint išskirti rūbus į nepriklausomus vienetus ir sumažinti triukšmą duomenyse. Atlikus segmentaciją, toliau galima pritaikyti kitus duomenų išgavimo metodus individualiems rūbams. Segmentacija taip pat gali būti panaudojama, norint išskirti ir kitas rūbų detales, pašalinti aplinkinį triukšmą: priežiūros aprašą, etiketę, drabužio žemesnio hierarchijos lygio klasifikaciją[32].

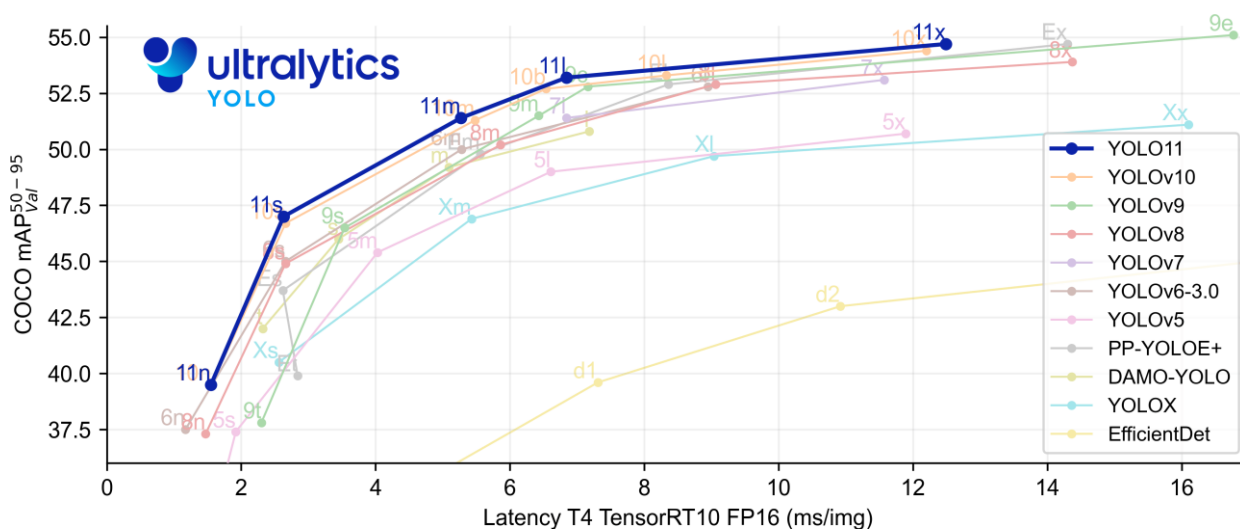
Objektų klasifikacijos metodai suteikia galimybę vaizdinėje medžiagoje atpažinti vaizdo tipą, pagal vaizdo savybes. Šiam uždaviniui spręsti paprastai naudojamas prižiūrimasis mašininis mokymas. Objektų klasifikacijos metodų panaudojimas drabužių skaitmeninio aprašo gavimo uždaviniui spręsti, suteiktų galimybę klasifikuoti segmentacijos metu išskirtus rūbus, taip pat atpažinti drabužio užsegimą, priežiūros aprašo iliustracijų reikšmes [33].

1.3. YOLO

YOLO (angl. You Only Look Once) architektūra buvo pristatyta 2015 metais. Šis modelis buvo sukurtas kaip objektams aptikti skirtas modelis. Šis modelis objektų aptikimo uždutį apibrėžia kaip regresijos problemą, ribinių langelių ir klasių prognozavimą sujungia į vieną bendrą iteraciją [34], skirtingai nei kiti modeliai, kurie objektų aptikimo uždutį atskiria į dvi dalis, kur pirmojoje dalyje vyksta regionų identifikavimas, o antrojoje jų klasifikavimas. Šis modelis padalina paveikslėlį į tinklėlį, kurio kiekvienas langelis atsakingas už objektų, kurių centrai patenka į jų ribojamą plotą aptikimą. Ši *YOLO* modelio architektūra leidžia vienu vertinimu nuspėti keletą ribinių langelių kartu su jų klasėmis.

YOLO modelį sudaro trys dalys: stuburas (angl. *backbone*), kaklas (angl. *neck*) ir galva (angl. *head*). Stuburo dalyje naudojamas gilusis konvoliucinis neuroninis tinklas, kuris yra skirtas savybėms išgauti iš įvesties. Kaklo dalis sujungia stuburą su galva – išgrynina erdvines savybes ir sujungia skirtingų gylių savybes. Šiame sluoksnyje gali būti naudojami papildomi konvoliucijų sluoksniai, savybių piramidės [35] ar kiti metodai, padedantys išgryninti savybes. Galvos dalis atsakinga už išvesties generavimą – šioje dalyje naudojami konvoliucijos sluoksniai kiekvieno *YOLO* išskaidyto tinklo langelio išvesčiai pateikti, į kurią įeina ribinių langelių koordinatės, objektingumo (*objectness*) įvertis ir klasių tikimybės [36].

YOLO architektūra yra nuolat tobulinama, išleidžiamos naujos šio modelio versijos. Kai kurios *YOLO* versijos palaiko ne vien objektų aptikimą, tačiau ir kitas užduotis kaip klasifikavimą, segmentaciją, žmogaus pozos nustatymą. Žemiau pateikiamas grafikas (žr. **8 pav.**), kaip kylant *YOLO* versijoms kyla modelio mAP50-95 metrikos.



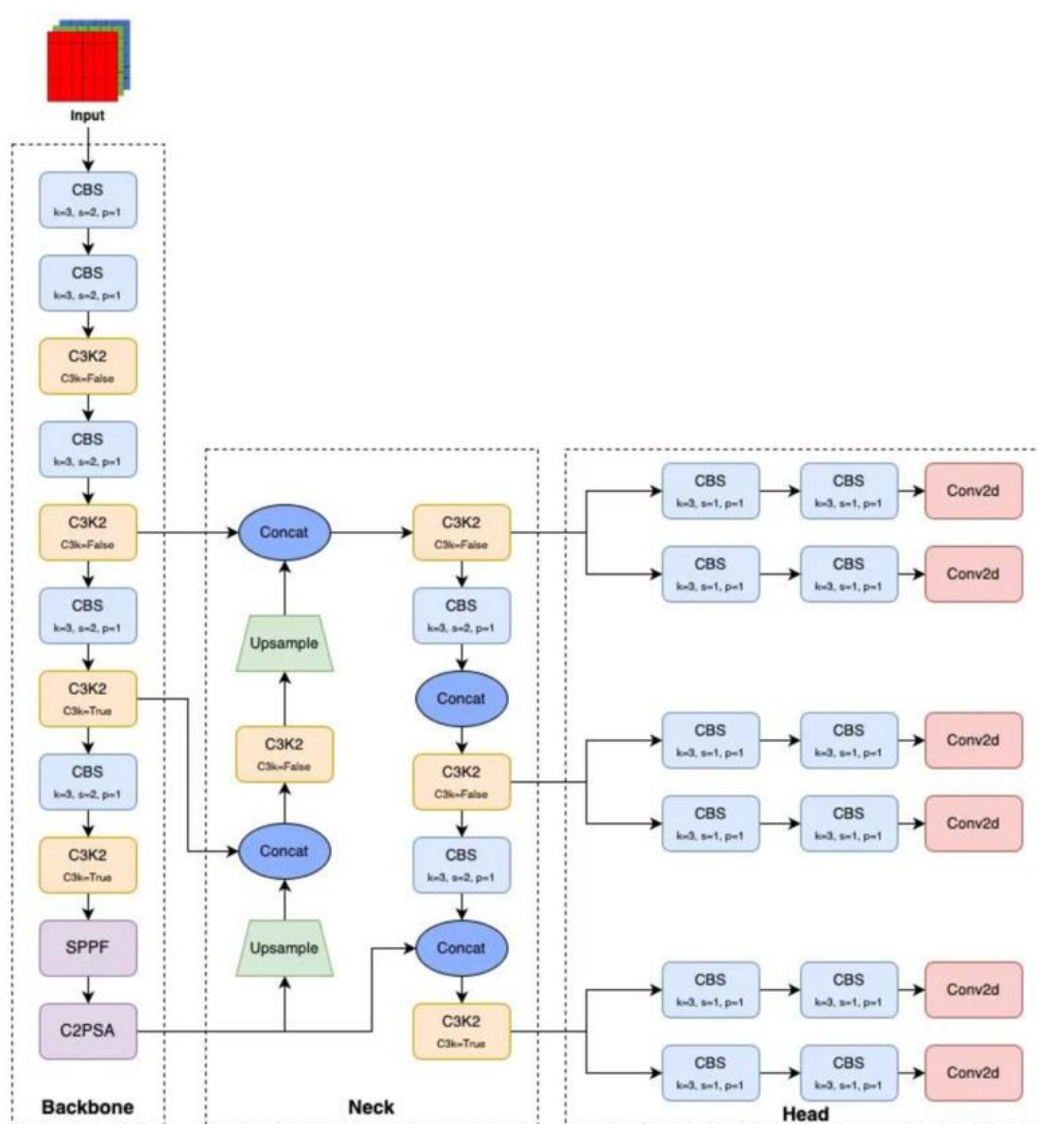
8 pav. YOLO modelio versijų ir jų skirtingų dydžių mAP50-95 metrikų palyginimas [37]

1.3.1. YOLO11

YOLO11 architektūra (žr. **9 pav.**) pristatyta 2024 metais. Ši architektūra remiasi ankstesnių *YOLO* modelių principais. Šioje modelio versijoje stuburo (*Backbone*) dalyje C2f bloką pakeičia C3k2 blokas, kuris naudoja dvi mažesnes konvoliucijų operacijas, vietoje vienos didelės konvoliucijos operacijos. Šis pokytis pagerina modelio greitaveiką. Modelyje naudojamas SPPF blokas, kuris buvo ir senesnėje modelio versijoje, tačiau po šio bloko pridamas C2SPA blokas, kuris leidžia geriau užfiksuoti erdvines savybes savybių žemėlapyje. Šis blokas leidžia modeliui koncentruotis į esminius paveikslėlio regionus, pagerinti objektų aptikimo tikslumą, kas yra naudinga kai aptinkami maži ar dalinai matomi objektai. Kaklo (*Neck*) dalyje C2F blokas kaip ir stuburo dalyje pakeistas C3k2 bloku. Galvos (*Head*) bloke naudojama keletas C3k2 blokų. Šie blokai naudojami keletose skirtingų galvos blokų vietų skirtinguose gyliuose. Galvos dalyje naudojami CBS blokai iškart po C3k2 blokų. Šie blokai išgrynina savybių žemėlapius, atlikdami normalizaciją, panaudodami SiLU aktyvacijos funkciją. Šie blokai atsakingi tiek už savybių išgavimą, tiek už objektų aptikimo procesą. Kiekviena objektų aptikimo atšaka užsibaigia Conv2D sluoksniais, kurie apjungia savybes į reikiamą kiekį išvesčių, reikalingų ribinio langelio (*bounding-box*) koordinatėms ir klasių išvesčiai [38].

Šis modelis pritaikomas daugeliui užduočių, susijusių su vaizdine informacija: objektams aptikti, segmentacijai, klasifikacijai, estimuoti žmogaus pozą, aptikti orientacinius objektus, objektams sekti.

Šio modelio variacija, skirta klasifikuoti vaizdus, kaip ir originali modelio versija, susideda iš stuburo ir kaklo dalies. Esminis skirtumas – šio modelio galvos dalis yra modifikuota ir vietoje aptikimo sluoksnio naudoja globalų vidurkinimą kartu su pilnai sujungtu klasifikavimo sluoksniu, kurio išvestis yra klasių tikimybės vektorius, vietoje standartiškai naudojamų ribinio langelio koordinatų, objektingumo (*objectness*) įverčių ir klasių įverčių kiekvienam aptiktam objektui.



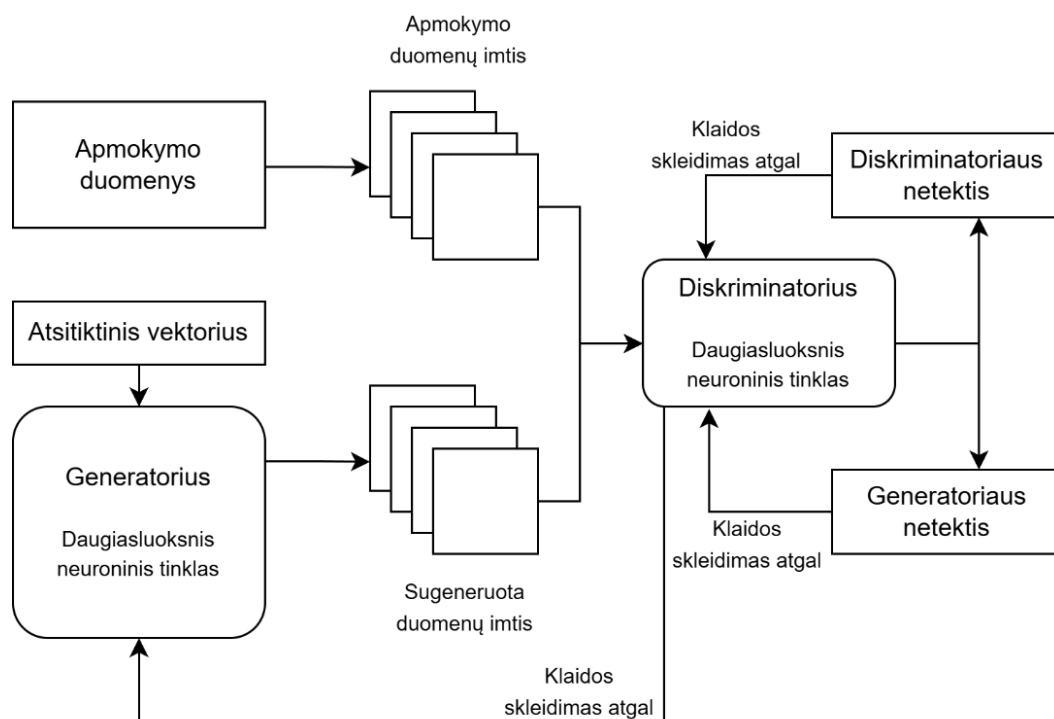
9 pav. YOLO11 architektūra [39]

1.4. GAN metodai

Generatyviniai priešiškieji neuroniniai tinklai (GAN) – mašininio mokymosi architektūros tipas, pristatytas 2014 metais [40]. Jį sudaro apjungti du neuroniniai tinklai: generatorius ir diskriminatorius. Generatorius – konvoliucinis neuroninis tinklas, transformuojantis atsitiktinį vektorių į sintetinius duomenis. Generatoriaus tikslas – sugeneruoti paveikslėlių, kurių diskriminatorius klasifikuoję kaip tikrą, atnaujinti savo svorius, siekti minimizuoti diskriminatoriaus tikslumą ir didinti jo praradimus.

Diskriminatorius – konvoliucinis neuroninis tinklas, apmokomas nustatyti ar sugeneruotas paveikslėlis yra tikras, ar sugeneruotas, jo tinklo svoriai atnaujinami siekiant minimizuoti klasifikavimo praradimus [40].

GAN metodo apmokymo seka (žr. **10 pav.**): sukuriama atsitiktinis standartinio skirstinio vektorius kiekvienam naujai generuojamam paveikslėliui, kuris yra perduodamas generatoriui. Generatorius panaudodamas triukšmo vektorius sukuria netikrą duomenį išvestį. Ši duomenų generavimo iteracija vyksta iki kol sugeneruojamas nustatytos imties dydžio generuotų duomenų rinkinys. Šis rinkinys kartu su apmokymo duomenų imtimi yra perduodama klasifikuoti diskriminatoriui. Diskriminatorius suklasifikuoja šiuos duomenis priskirdamas tikimybės reikšmę [0-1] intervale – didesnė reikšmė nurodo didesnę tikimybę, kad paveikslėlis yra tikras. Diskriminatoriui suklasifikavus duomenų imtį vyksta diskriminatoriaus ir generatoriaus klaidos skleidimas atgal. Diskriminatoriaus svoriai yra atnaujinami pagal tai, kaip teisingai jis suklasifikuoja tikrą ar sugeneruotą paveikslėlį. Generatorius svoriai yra atnaujinami pagal tai, kaip jam pavyko apgauti diskriminatorių, sugeneruojant kuo realistiškesnį paveikslėlį, gaunant tikimybę, kuri yra kuo artimesnė 1. Šis procesas kartojamas iki kol modelis apsimoko pagal nustatytus baigties parametrus.



10 pav. GAN architektūra

1.4.1. *StyleGAN*

StyleGAN architektūra pristatyta 2018 metais [41]. Tai yra patobulinta GAN architektūra, kurioje modifikuota generatoriaus architektūra – panaudojamas stiliaus perkėlimas, pritaikant *AdaIN* normalizaciją ir naudojant progresyvią augimo techniką, taip išgaunant aukštesnės kokybės ir realistiškesnius paveikslėlius. Apmokymo metu paveikslėliai pirmiausiai generuojami žemos kokybės, o su kiekviena kita iteracija jų kokybė didėja.

StyleGAN skiriasi nuo kitų GAN metodų savo generatoriumi (žr. **11 pav.**). Vietoj to, kad panaudotų atsitiktinį standartinio skirstinio vektorius (z) tiesiogiai generuoti paveikslėlį, metodas vektorius

AdaIN – normalizacijos metodas, įprastai taikomas stiliaus perkėlimo užduotims, turinio duomenų ir stiliaus duomenų vidurkio ir dispersijos suderinimui, taip pritaikydamas stilių turinio duomenims [42]. Skaičiavimas atliekamas pagal (1) formulę:

$$AdaIN(x, y) = \sigma(y) \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y) \quad (1)$$

čia $\sigma(y)$ ir $\mu(y)$ – turinio savybių vidurkis ir dispersija, $\sigma(x)$ ir $\mu(x)$ – stiliaus savybių (transformuoto tarpinio vektoriaus) vidurkis ir dispersija.

1.4.2. *StyleGAN2*

StyleGAN2 architektūra pristatyta 2020 metais, kuri yra pagerinta *StyleGAN* versija [43]. Šioje versijoje išliko MLP panaudojimas tarpinio vektoriaus sukūrimui, stiliaus pritaikymas sintezės dalyje. Pašalintas *AdaIN* normalizacijos metodas – pakeistas svorio moduliacijos metodu ir demoduliacija. Moduliacijos metu konvoliucijos svoriams pritaikomas transformuotas stiliaus (w) vektorius. Demoduliacijos metu, moduliacijos išvestis yra normalizuojama. Naujojoje architektūroje taip pat nebe naudojamas progresinis augimas – neuroninis tinklas apmokomas panaudojant pilną ir nebeaugantį generatoriaus ir diskriminatoriaus neuroninį tinklą, dėl ko gaunami stabilesni išvesties rezultatai. Šie pakeitimai nulėmė realistiškesnę, aukštesnės kokybės paveikslėlių generavimą ir modelio patikimumą, lyginant su buvusiu modelio versija.

1.4.3. *StyleGAN2-ADA*

StyleGAN2-ADA architektūra pristatyta 2020 metais. Tai yra *StyleGAN2* architektūros variacija, kurioje pritaikyta adaptivi duomenų augmentacija (ADA), kuri stabilizuoja diskriminatoriaus apmokymą, kai apmokymui naudojamas mažas duomenų rinkinys, kuriame yra mažiau nei 30 tūkst. paveikslėlių [44]. Šis metodas yra tinkamas pasirinkimas apmokyti mažos apimties duomenų rinkinius, kadangi geba sugeneruoti aukštos kokybės vaizdus, turint limituotą kiekį duomenų [45], [46], [47], [48]. Augmentacija pritaikoma tiek tikriems, tiek sugeneruotiems paveikslėliams, prieš perduodant juos klasifikuoti diskriminatoriui. Augmentacijos funkcijos yra pritaikomos atsižvelgiant į diskriminatoriaus išvestį – jei diskriminatorius atpažįsta tikrus ir sugeneruotus paveikslėlius aukštu pasitikėjimu – diskriminatorius priskiria paveikslėliams reikšmes, stipriai atitolusias nuo nulio, tada augmentacijos funkcijos yra stiprinamos, siekiant sumažinti diskriminatoriaus spėjimo pasitikėjimą ir artinti išvestį prie 0. Jeigu diskriminatoriaus išvestis artėja link 0, augmentacijos yra silpninamos arba išlaikomos tokios pat.

1.4.4. *StyleGAN3*

StyleGAN3 architektūra pristatyta 2021 metais. Ši architektūra yra patobulinta *StyleGAN2* architektūra. Pagrindiniai pakeitimai buvo atlikti generatoriaus dalyje. Pašalintas triukšmas įnešamas tarp sluoksnių ir absoliučių koordinatų naudojimas, dėl kurio paveikslėliuose atsirasdavo realybės neatitinkančių artefaktų (tekstūros neatitikimų, kai objektas paveikslėlyje pakeičia poziciją, deformacijų, vaizdo dalių nenatūralaus ryškumo). Šioje architektūroje vietoje generatoriaus pirmojo sluoksnio įvesties, kuri anksčiau buvo konstanta, dabar naudojamos Furjė savybės (*Fourier features*), šios savybės yra duomenys apie paveikslėlio erdvinę informaciją, kurios naudojamos pirminiam savybių žemėlapiui sukurti. Furjė savybių įvedimas leidžia pašalinti realybės neatitinkančių artefaktų generavimą ir padidinti erdvės pokyčių generavimo realistiškumą. Rezoliucijos didinimui

nebenaudojamas artimiausių kaimynų metodas. Vietoj to, prieš padidinant rezoliuciją, pritaikomas filtras, kuris pašalina didelio intensyvumo komponentus, kurie gali iškraipyti paveikslėlį. Šie pakeitimai leidžia *StyleGAN3* generuoti realistiškesnius paveikslėlius, išlaikant erdvės vientisumą.

2. Sprendimo projektavimas mados vaizdų atpažinimui

2.1. Reikalavimų specifikacija

2.1.1. Sistemos paskirtis

Sistema yra skirta drabužių duomenų išgavimui iš nuotraukos, pritaikant giliuosius neuroninius tinklus. Sistemos įvestis – nuotrauka, kurioje pavaizduotas drabužis. Sistema turi gebėti išgauti reikiamą informaciją iš drabužių nuotraukų. Sistemos išvestis – drabužio duomenys, išgauti iš nuotraukos.

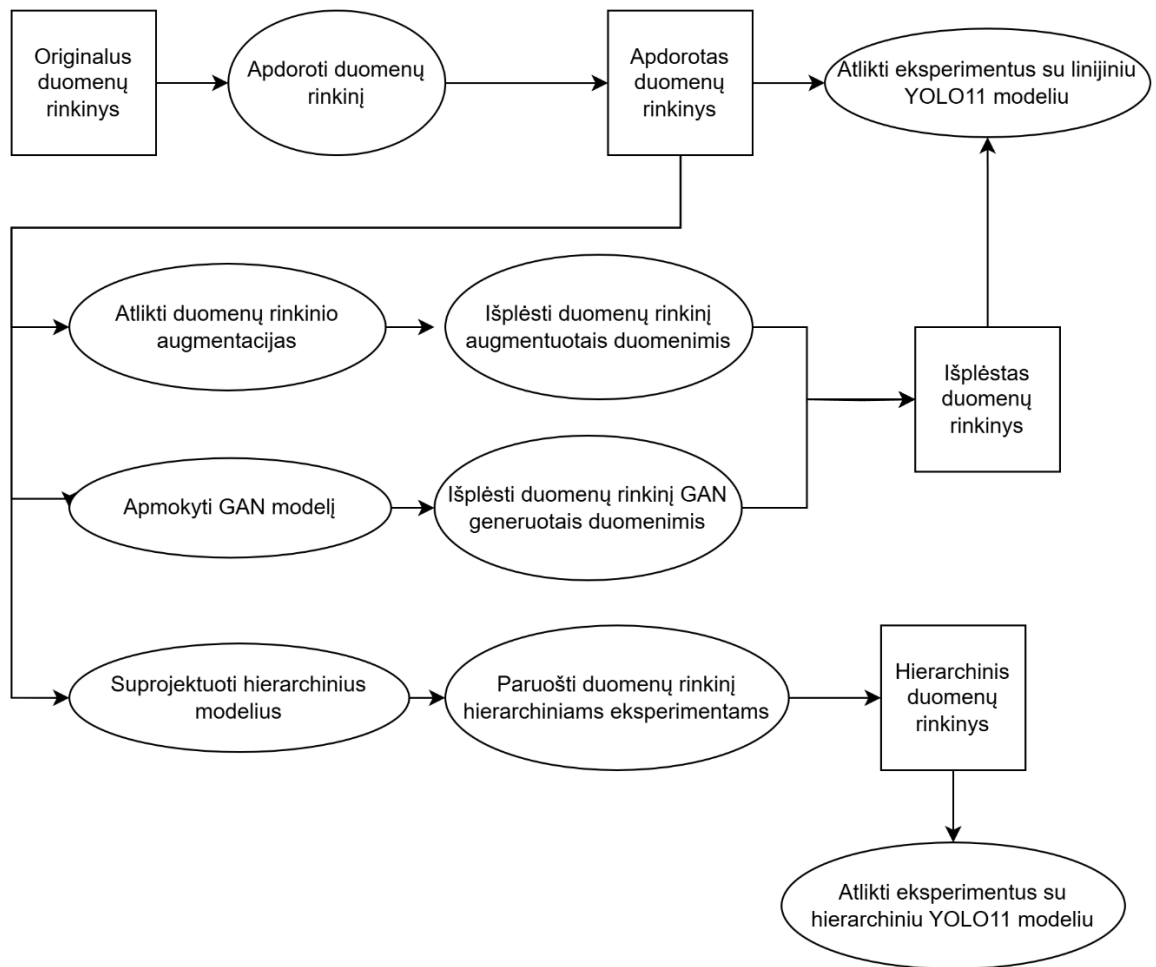
Sukurta sistema gali palengvinti drabužių duomenų skaitmenizavimą, keliant jų informaciją į internetinę erdvę. Ši sistema gali būti aktuali drabužių perdirbimo, pardavimo srityse ar kitose srityse kuriose reikalingas informacijos iš drabužių išgavimo automatizavimas.

2.1.2. Tyrimo planas

Tyrimas susideda iš kelių dalių:

1. Apdoroti ir paruošti eksperimentams skirtą duomenų rinkinį.
2. Suprojektuoti hierarchinius modelius.
3. Apmokyti GAN modelius ir sukurti praplėstus apmokymo duomenų rinkinius.
4. Apmokyti linijinius modelius su paruoštu duomenų rinkiniu.
5. Apmokyti hierarchinius modelius su paruoštu duomenų rinkiniu.
6. Apmokyti linijinius modelius su praplėstais duomenų rinkiniais.
7. Atlikti eksperimentus ir apskaičiuoti modelių tikslumo metrikas.
8. Palyginti rezultatus ir pateikti išvadas.

Šio tyrimo planas ir eiliškumas pavaizduoti žemiau pateiktame paveikslėlyje (žr. **12 pav.**).



12 pav. Tyrimo schema

2.1.3. Projekto kūrimo įrankiai ir aplinka

Projekto bandymams bus naudojama aplinka su šiais parametrais:

- AMD EPYC 7452 32-Core Processor
- 256GB RAM
- GPU NVIDIA A100-PCIE-40GB

Sistemos kūrimui naudojami įrankiai:

1. Naudojama programavimo kalba – *Python*, 3.7, 3.8 ir 3.11 versijos
2. Paveikslėlių generavimui naudojami *StyleGAN2-ADA* ir *StyleGAN3* modeliai.
3. Vaizdinės medžiagos duomenų išgavimo tyrimui naudojamas *YOLOv11-cls* modelis.

Pasirinkta *Python* programavimo kalba, kadangi tai yra populiariausia kalba, naudojama mašininio mokymosi modeliams kurti, turinti didelį kiekį tam skirtų ir nuolat atnaujinamų bibliotekų. Taip pat ši kalba turi plačią naudotojų bendruomenę, todėl rasti informaciją ar atsakymus apie iškilusius klausimus yra paprasta.

2.1.4. Funkciniai reikalavimai

Sistemos funkciniai reikalavimai:

1. Sistema turi palaikyti *.jpg* paveikslėlių formatą.

2. Vienu metu gali būti įkeliama tik 1 nuotrauka, ne didesnė nei 2MB.

Norint išvengti skirtingų formatų paveikslėlių apdorojimo įgyvendinimo ir su juo susijusių problemų, nuspręsta palaikyti vieno formato nuotraukų tipą. Sistema palaikys vienos nuotraukos ir iki 2MB dydžio įkėlimą, siekiant išvengti sistemos resursų problemų.

2.1.5. Vertinimo metrikos

Tyrime bus naudojamos *top-k* metrikos rezultatų palyginimui su egzistuojančiais sprendimais, naudotais pasirinktam duomenų rinkiniui. Apmokyti modeliai taip pat bus vertinami naudojant tikslumo (angl. *Accuracy*), preciziškumo (angl. *Precision*), jautrumo (angl. *Recall*), *F1* ir jų svorines metrikas.

Top-k tikslumas (2) – metrika, naudojama multi-klasių klasifikavimo modeliams vertinti, nurodanti kokia dalis tikrųjų klasių yra tarp *k* aukščiausios tikimybės modelio prognozuojamų klasių, apskaičiuojama pagal formulę:

$$Top_k = \frac{1}{N} \sum_{i=1}^N 1(y_i \in \hat{Y}_i^k) \quad (2)$$

čia *N* – bendras klasifikuojamos imties dydžio kiekis, y_i – tikroji imties vieneto klasė, \hat{Y}_i^k – *k* dydžio aibė, kurioje yra *k* aukščiausios tikimybės prognozių *i* duomenų vienetai. $1(\cdot)$ – indikatoriaus funkcija, kuri lygi 1, jei y_i reikšmė yra \hat{Y}_i^k aibėje ir 0, jei reikšmės šioje aibėje nėra.

Tikslumas – metrika, naudojama nurodyti bendrą modelio tikslumą, parodanti teisingai suklasifikuotų klasių kiekį santykiu su visų prognozių kiekiu, apskaičiuojama pagal formulę:

$$Tikslumas = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

čia *TP* – teisingų teigiamų prognozių kiekis, *TN* – teisingų neigiamų prognozių kiekis, *FP* – neteisingų teigiamų prognozių kiekis, *FN* – neteisingų neigiamų prognozių kiekis.

Preciziškumas – metrika, nurodanti kokia dalis teisingų teigiamų prognozių yra prognozuojama teisingai, vengiant neteisingų teigiamų prognozių. Aukštas preciziškumas rodo didesnę modelio patikimumą, kuris yra svarbus tais atvejais, kai klaidingas teigiamas klasifikavimas nėra toleruotinas. Preciziškumas apskaičiuojamas pagal formulę:

$$Preciziškumas = \frac{TP}{TP + FP} \quad (4)$$

čia *TP* – teisingų teigiamų prognozių kiekis, *FP* – neteisingų teigiamų prognozių kiekis.

Jautrumas – metrika, naudojama nurodyti kokia dalis teigiamai klasifikuotų atvejų yra prognozuojama teisingai, vengiant neteisingų neigiamų prognozių. Aukštas jautrumas rodo didesnę modelio patikimumą tais atvejais, kai svarbiau prognozuoti teigiamą klasę, toleruojant klaidingą teigiamą prognozę. Jautrumas apskaičiuojamas pagal formulę:

$$Jautrumas = \frac{TP}{TP + FN} \quad (5)$$

čia TP – teisingų teigiamų prognozių kiekis, FN – neteisingų neigiamų prognozių kiekis.

F1 – metrika, nurodanti bendrą santykį tarp preciziškumo ir jautrumo. Naudojama kai svarbus tiek klaidingas teigiamas klasifikavimas, tiek klaidingas neigiamas klasifikavimas. *F1* apskaičiuojamas pagal formulę:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (6)$$

čia P – preciziškumas pagal formulę (4), R – jautrumas pagal formulę (5).

Svorinis tikslumas, preciziškumas, jautrumas, *F1* – metrika, naudojama, kai duomenų rinkinys nėra subalansuotas ir klasių balansas turi atsispindėti metrikose – didesnių klasių klasifikavimo metrikos yra svarbesnės. Svorinės metrikos skaičiuojamos pagal formulę:

$$X = \frac{1}{N} \sum_{i=1}^N w_i \cdot X_i \quad (7)$$

čia N – klasių kiekis, w_i – i klasės svoris, X_i – i klasės tikslumas pagal formulę (3), preciziškumas pagal formulę (4), jautrumas pagal formulę (5) arba *F1* pagal formulę (6).

2.1.6. Nefunkciniai reikalavimai

Sistemos nefunkciniai reikalavimai:

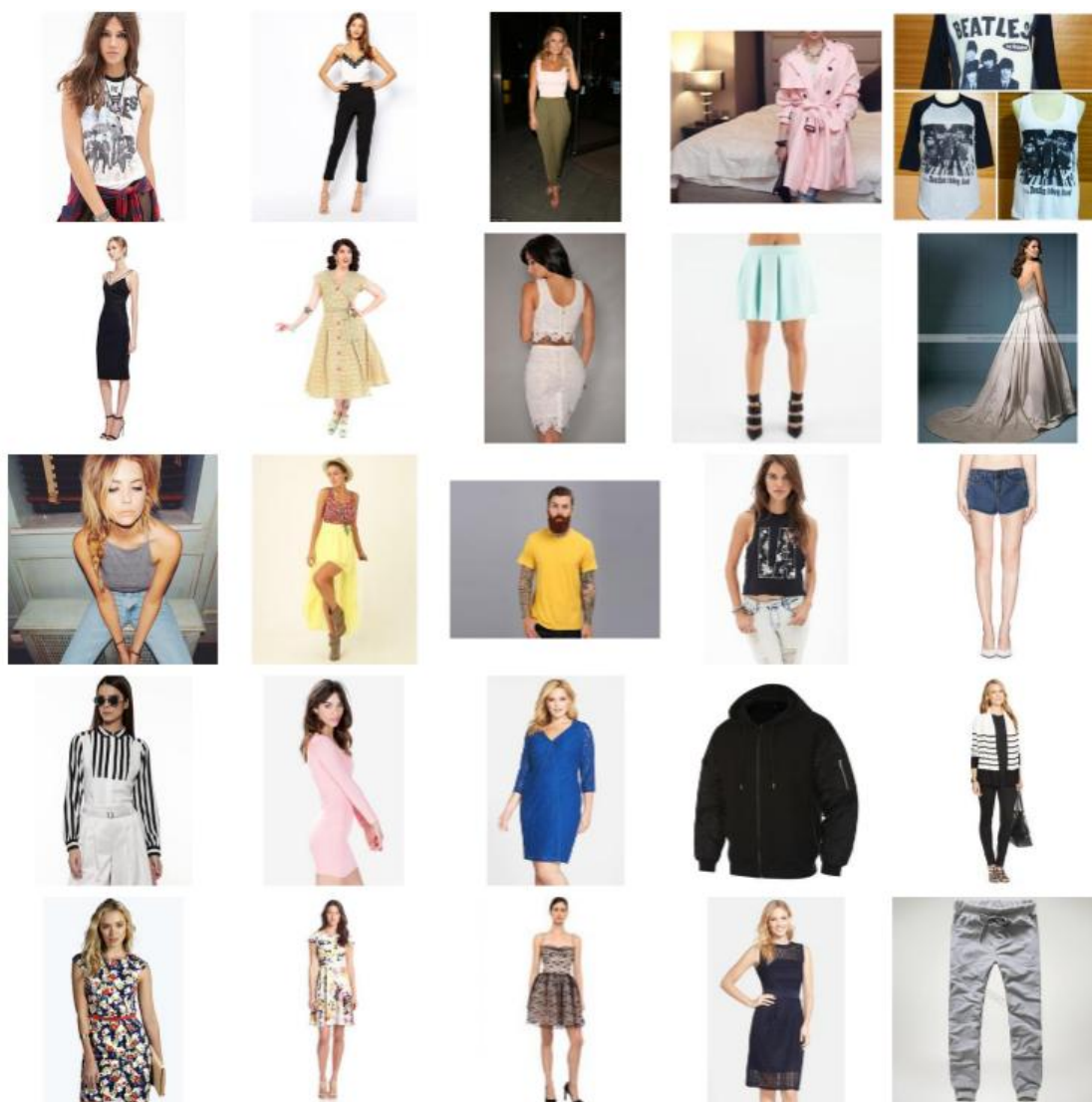
1. Vieno prognozavimo trukmė turėtų būti ne ilgesnė nei 1s.
2. Modelio tikslumas turi pasiekti iš anksto numatytas slenkstines vertes – *top-3* turi būti ne mažesnis nei 91.99%, *top-5* – 96.44%.

2.2. Duomenų rinkinys

Šiame darbe naudojamas *DeepFashion* [1] duomenų rinkinys, kurį sudaro virš 800 tūkst. drabužių paveikslėlių. Paveikslėliams priskirti 1 tūkst. atributų reikšmių ir kiekvienas jų turi aprašytas orientacinių taškų koordinatas. Šis rinkinys šiuo metu yra didžiausias viešai prieinamas mados vaizdų duomenų rinkinys. Tyrime bus naudojami duomenų rinkinio dalis, skirta kategoriniam klasifikavimui.

Kategorinių duomenų rinkinį sudaro 50 kategorijų, kurių 46 kategorijos turi priskirtus paveikslėlius. Šiame rinkinyje yra 289,222 paveikslėliai (žr. **1 lentelė**). Domenų rinkinyje 209,222 (72.40%) nuotraukų priskirtos apmokymo duomenų rinkiniui, 40,000 (13.8%) validacijos duomenų rinkiniui ir 40,000 (13.8%) nuotraukų testavimui. Nors duomenų rinkinį sudaro didelis kiekis paveikslėlių, tačiau net 24 kategorijos neturi daugiau nei 1,000 paveikslėlių, o 9 jų ne daugiau nei 100. Tačiau didžiosios kategorijos turi keliasdešimt tūkstančių paveikslėlių. Duomenų rinkinys yra nesubalansuotas, dėl ko šio rinkinio klasifikavimas yra komplikuoja užduotis.

Duomenų rinkinio sudarymui panaudotos nuotraukos iš internetinių drabužių parduotuvių. Kiekvienas drabužis užfiksuotas 4 arba 5 skirtingais kampais. Duomenų rinkinį sudaro tiek parduotuvių drabužių nuotraukos, tiek žmonių užfiksuotos drabužių nuotraukos (žr. **13 pav.**).

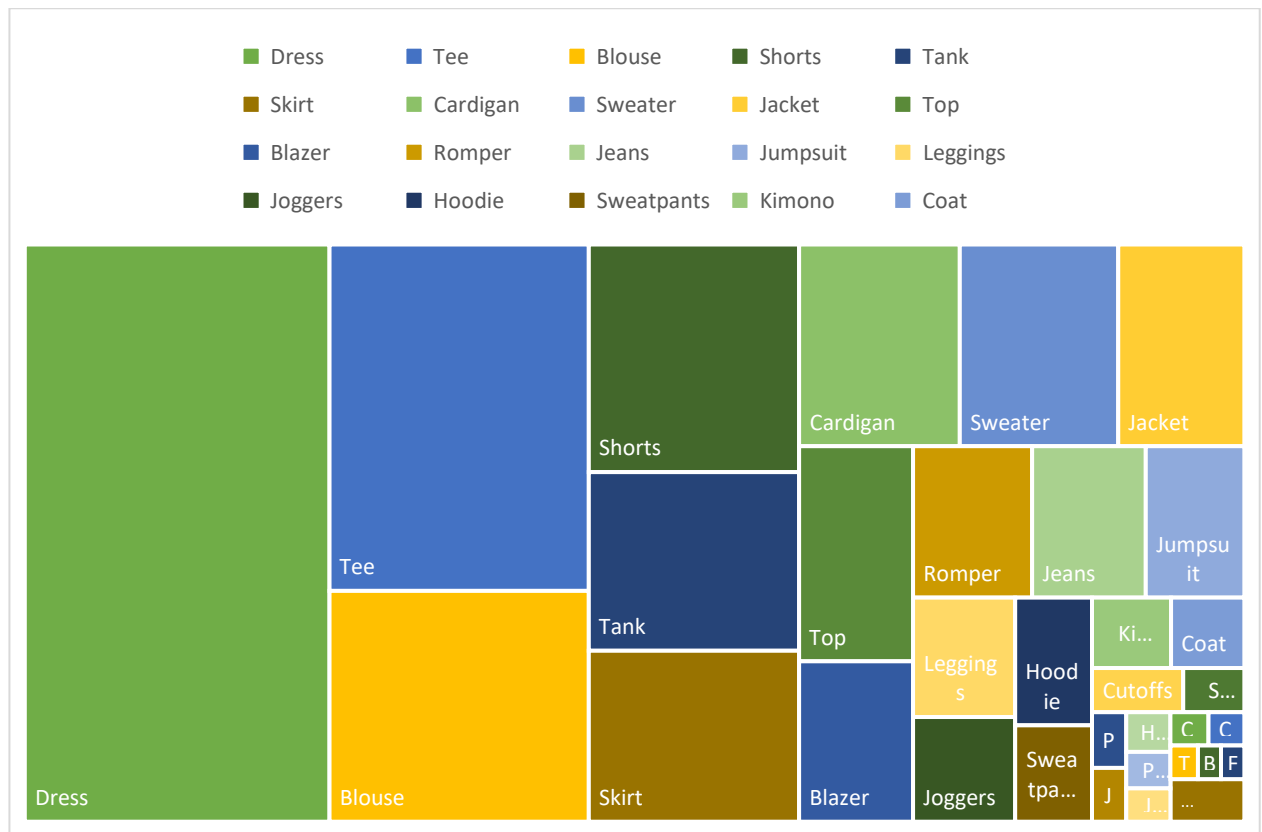


13 pav. *DeepFashion* duomenų rinkinio paveikslėlių pavyzdžiai [1]

1 lentelė. Duomenų rinkinio klasės, jų kiekis ir procentinė dalis

Kategorija	Kiekis duomenų rinkinyje	Duomenų rinkinio dalis (%)
<i>Dress</i>	72158	24.95
<i>Tee</i>	36887	12.75
<i>Blouse</i>	24557	8.49
<i>Shorts</i>	19666	6.79
<i>Tank</i>	15429	5.33
<i>Skirt</i>	14773	5.10
<i>Cardigan</i>	13311	4.60
<i>Sweater</i>	13123	4.53
<i>Jacket</i>	10467	3.61
<i>Top</i>	10078	3.48

<i>Blazer</i>	7495	2.59
<i>Romper</i>	7408	2.56
<i>Jeans</i>	7076	2.44
<i>Jumpsuit</i>	6153	2.12
<i>Leggings</i>	5013	1.73
<i>Joggers</i>	4416	1.52
<i>Hoodie</i>	4048	1.39
<i>Sweatpants</i>	3048	1.05
<i>Kimono</i>	2294	0.79
<i>Coat</i>	2120	0.73
<i>Cutoffs</i>	1669	0.57
<i>Sweatshorts</i>	1106	0.38
<i>Poncho</i>	791	0.27
<i>Jersey</i>	748	0.26
<i>Henley</i>	716	0.25
<i>Parka</i>	676	0.23
<i>Jeggings</i>	594	0.21
<i>Chinos</i>	527	0.18
<i>Culottes</i>	486	0.17
<i>Trunks</i>	386	0.13
<i>Button-Down</i>	330	0.11
<i>Flannel</i>	324	0.11
<i>Bomber</i>	309	0.10
<i>Anorak</i>	160	0.05
<i>Robe</i>	150	0.05
<i>Turtleneck</i>	146	0.05
<i>Kaftan</i>	126	0.04
<i>Peacoat</i>	97	0.03
<i>Capris</i>	77	0.03
<i>Onesie</i>	70	0.02
<i>Caftan</i>	54	0.02
<i>Gauchos</i>	49	0.02
<i>Jodhpurs</i>	45	0.02
<i>Sarong</i>	32	0.01
<i>Halter</i>	17	0.01
<i>Coverup</i>	17	0.01



14 pav. Duomenų rinkinio klasių pasiskirstymas

2.2.1. Duomenų rinkinio paruošimas

Atlikta duomenų rinkinio analizė parodė, kad duomenų rinkinys yra stipriai išbalansuotas. Didžioji dalis duomenų – 24.94% yra *Dress* kategorijos paveiksliai. Antroji didžiausia kategorija – *Tee*, su 12.75% paveikslėlių. Kitos kategorijos sudaro tik po kelis procentus duomenų rinkinio. Buvo nuspręsta iš duomenų rinkinio pašalinti klases, kurios sudaro mažiau nei 0.3% duomenų rinkinio, jas apjungiant į didesnes klases, jei šios klasės yra artimos. Išmestos klasės: *Poncho*, *Robe*, *Culottes*, *Anorak*, *Kaftan*, *Onesie*, *Caftan*, *Gauchos*, *Jodhpurs*, *Coverup*. Apjungtos klasės:

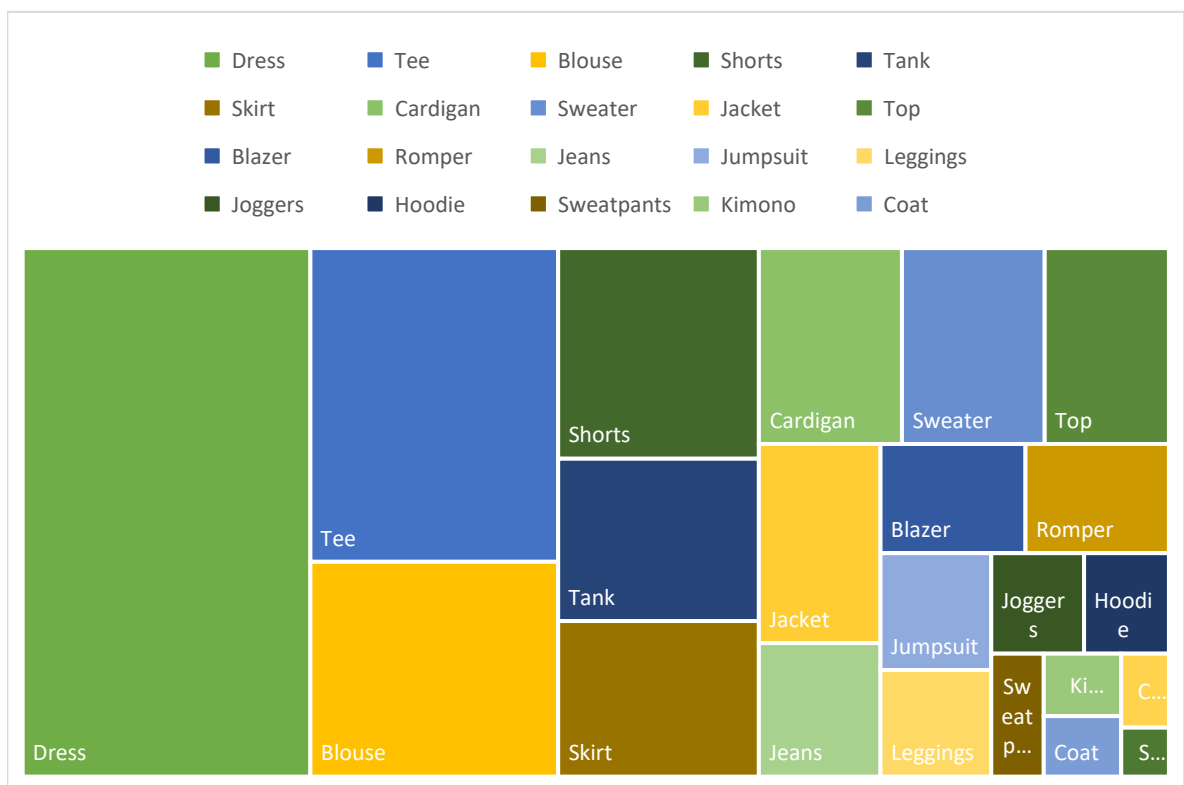
- Į *Top* klasę: *Jersey*, *Henley*
- Į *Jacket* klasę: *Parka*, *Bomber*
- Į *Leggings* klasę: *Jeggings*
- Į *Jeans* klasę: *Chinos*, *Capris*
- Į *Shorts* klasę: *Trunks*,
- Į *Blouse* klasę: *Button-down*, *Flanel*
- Į *Sweater* klasę: *Turtleneck*
- Į *Coat* klasę: *Peacoat*
- Į *Skirt* klasę: *Sarong*
- Į *Tank* klasę: *Halter*

Apjungus šias klases gaunamas duomenų rinkinys (žr. 2 lentelė), kurio klasių pasiskirstymas pateiktas 15 pav..

2 lentelė. Apjungtų klasių duomenų rinkinio klasės ir jų procentinė dalis

Kategorija	Kiekis duomenų rinkinyje	Duomenų rinkinio dalis (%)
<i>Dress</i>	72158	25.12

<i>Tee</i>	36887	12.84
<i>Blouse</i>	25211	8.78
<i>Shorts</i>	20052	6.98
<i>Tank</i>	15446	5.38
<i>Skirt</i>	14805	5.15
<i>Cardigan</i>	13311	4.63
<i>Sweater</i>	13269	4.62
<i>Jacket</i>	11452	3.99
<i>Top</i>	11542	4.02
<i>Blazer</i>	7495	2.61
<i>Romper</i>	7408	2.58
<i>Jeans</i>	7680	2.67
<i>Jumpsuit</i>	6153	2.14
<i>Leggings</i>	5607	1.95
<i>Joggers</i>	4416	1.54
<i>Hoodie</i>	4048	1.41
<i>Sweatpants</i>	3048	1.06
<i>Kimono</i>	2294	0.80
<i>Coat</i>	2217	0.77
<i>Cutoffs</i>	1669	0.58
<i>Sweatshorts</i>	1106	0.38



15 pav. Apjungto duomenų rinkinio klasių pasiskirstymas

Kadangi duomenų rinkinį sudaro paveikslėliai, kuriuose gali pasitaikyti nuotraukų, kuriose yra keli drabužiai vienoje nuotraukoje, paveikslėliai buvo apkarpyti, paliekant paveikslėlyje tik tą drabužį, kuris turi būti suklasifikuotas. Tai buvo atlikta, panaudojant duomenų rinkinio ribinių langelių (*bounding box*) duomenis. Kiekvienas iškirptas vaizdas buvo standartizuojamas, kad ilgoji paveikslėlio kraštinė prilygtų 256 pikseliams, prie trumposios kraštinės pridėdant baltą foną, kad abi paveikslėlio kraštinės būtų 256 pikselių ilgio.

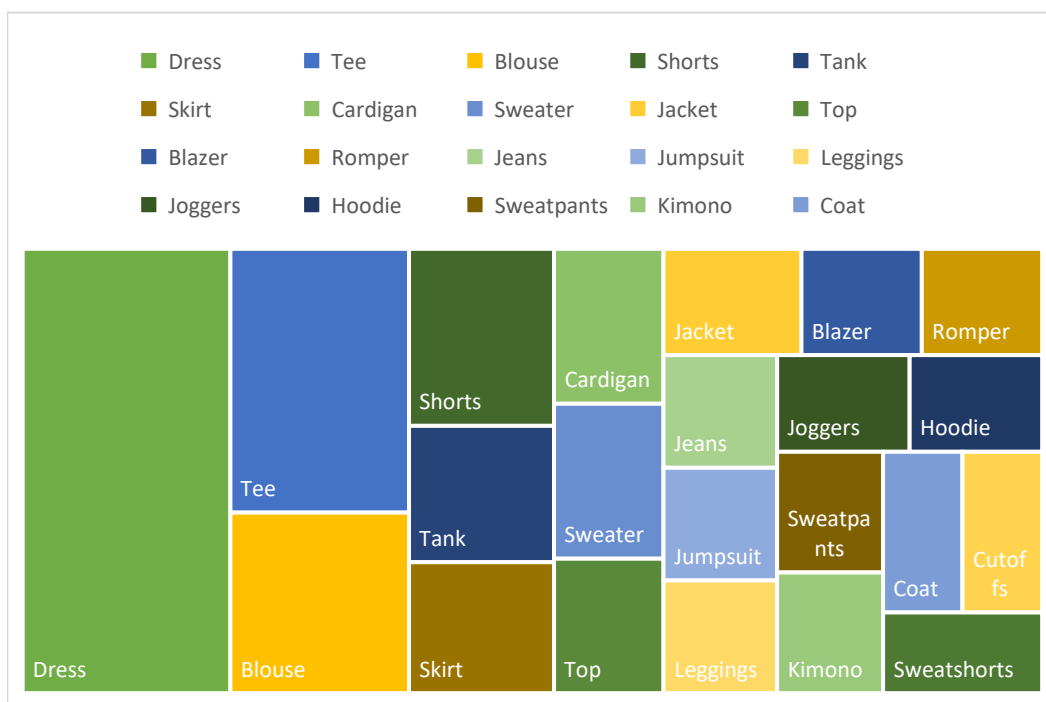
Iš šio duomenų rinkinio paruošta keletas variacijų, hierarchiniam klasifikavimui:

1. Trijų lygių hierarchijos duomenų rinkinys
2. Dviejų lygių hierarchijos duomenų rinkinys
3. Trijų nepilnų lygių hierarchijos duomenų rinkinys

Šie duomenų rinkiniai naudojami skirtingų hierarchijų tipų modeliams apmokyti.

Be duomenų paruošimo hierarchiniam klasifikavimui, buvo sugeneruoti duomenų rinkiniai klasėms balansuoti. Apmokymo duomenų rinkiniui buvo pritaikyta augmentacijos ir GAN metodai, kiekvienu atveju papildant mažesnes nei 10,000 paveikslėlių klases iki 7,240 (72,40% bendro duomenų rinkinio dalies) apmokymui skirtų paveikslėlių.

Pritaikius šias metodikas duomenų rinkiniui išplėsti, apmokymo duomenų rinkinys tapo labiau subalansuotas (žr. **16 pav.**). Klasės, kurios prieš tai buvo nepakankamai reprezentatyvios ir sudarė mažiau nei 3.48% duomenų rinkinio, tapo tolygiau pasiskirsčiusios. Didžiosios klasės taip pat tapo mažiau dominuojančiomis: *Dress* (prieš tai 25.12%, dabar 20.38%), *Tee* (prieš tai 12.84%, dabar 10.42%) klasės, kurios prieš tai sudarė didžiąją dalį duomenų rinkinio, dabar sudaro mažesnę dalį. Mažiausios klasės dabar apima po 2.82% duomenų rinkinio, kai prieš tai buvo intervale nuo 0.38%-2.61%. Šiais duomenų rinkinio pakeitimais siekiama padidinti mažųjų klasių reprezentatyvumą, giliojo mokymosi modelių apmokymo metu, išvengti modelio šališkumo didžiųjų klasių atžvilgiu.



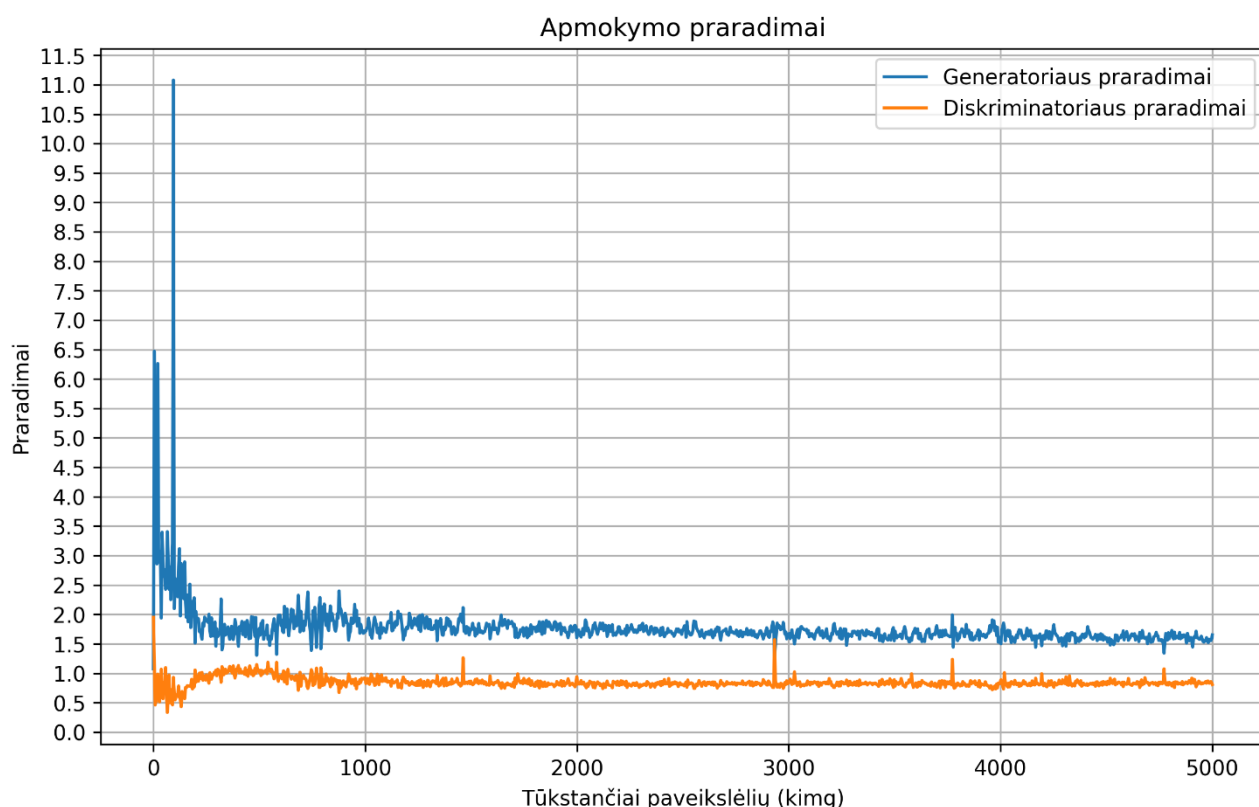
16 pav. Išplėsto duomenų rinkinio klasių pasiskirstymas

2.2.2. GAN modelių apmokymas duomenų sintezei

Apmokymo duomeniui išplėsti, panaudoti *StyleGAN2-ADA* ir *StyleGAN3* modeliai. Modeliai apmokyti panaudojant 5,000,000 paveikslėlių kiekvienai klasei.

2.2.2.1. *StyleGAN2-ADA*

Modelis apmokytas panaudojant *paper256* konfigūraciją, skirtą 256x256 px paveikslėliams, 128 imties (*batch*) dydį. Mokymo metu, generatoriaus ir diskriminatoriaus praradimai stabilizavosi, apdorojus maždaug du milijonus paveikslėlių kiekvienai klasei (žr. **17 pav.**). Generatoriaus praradimai iteracijų pradžioje labai dideli (6.5-11), tačiau pamažu stabilizuojasi ties 1.5-2 reikšme. Diskriminatoriaus praradimai iteracijų pradžioje taip pat didesni, tačiau greitai stabilizuojasi ties 0.5-1 reikšme.



17 pav. *StyleGAN2-ADA* generatoriaus ir diskriminatoriaus praradimai apmokymo metu (*Blazer* klasė)

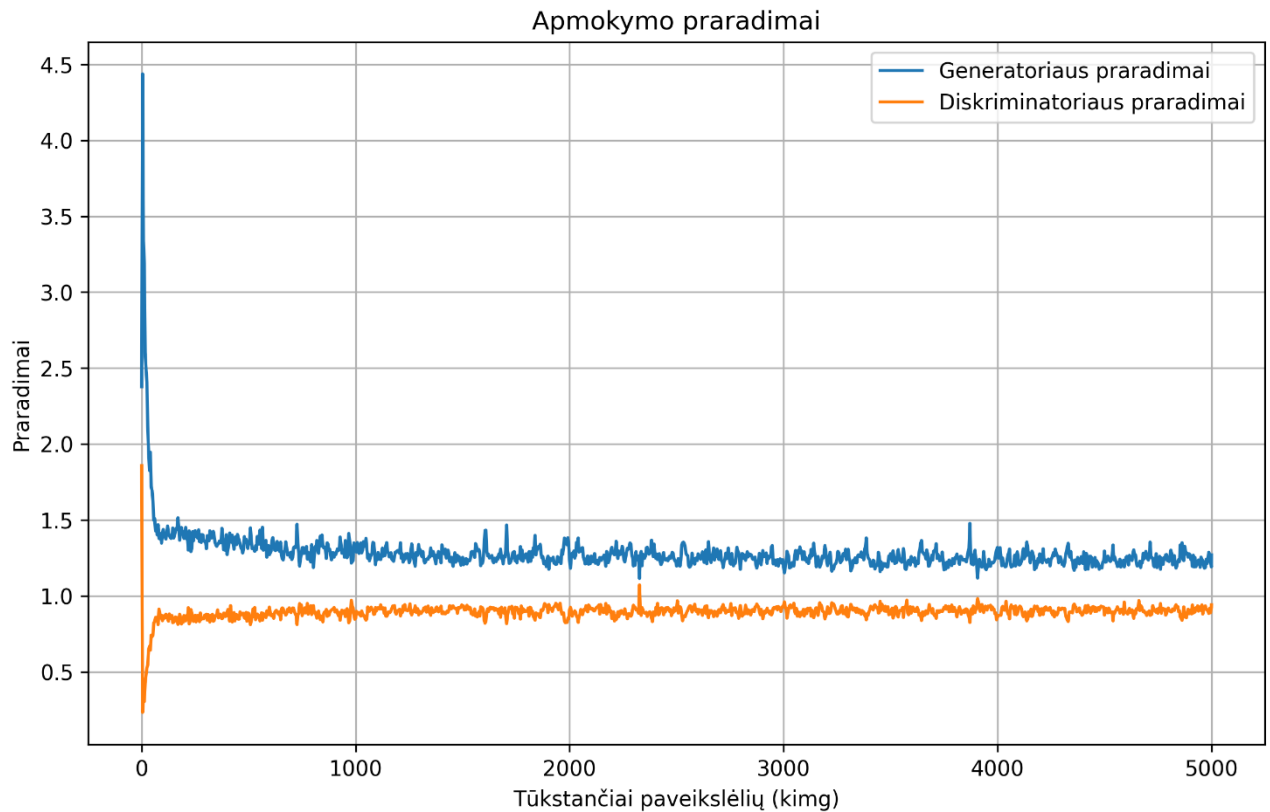
Apmokius modelius, generuojami paveikslėliai duomenų rinkinio papildymui. Žemiau pateikti *Blazer* klasės generuoti paveikslėlių pavyzdžiai (žr. **18 pav.**). Generuotuose paveikslėliuose vizualiai matosi klasės savybės, iš kurių galima spręsti kad tai yra *Blazer* klasės paveikslėlis, tačiau galima matyti, kad tai yra generuotas paveikslėlis iš tam tikrų nenatūralių artefaktų, paliktų paveikslėliuose.



18 pav. *StyleGAN2-ADA Blazer* klasės sugeneruotų paveikslėlių pavyzdys

2.2.2.2. *StyleGAN3*

Modelis apmokytas panaudojant *stylegan3-r-ffhq-256x256* konfigūraciją, skirtą 256x256 px paveikslėliams, 64 imties (*batch*) dydį. Mokymo metu, generatoriaus ir diskriminatoriaus praradimai stabilizavosi, apdorojus maždaug milijoną paveikslėlių kiekvienai klasei (žr. **19 pav.**). Generatoriaus praradimai iteracijų pradžioje dideli (2.5-4.5), tačiau daugiau nei du kartus mažesni nei *StyleGAN2-ADA* modelio. Jie pamažu stabilizuojasi ties 1-1.5 reikšme, taip pat žemesne nei senesnė modelio versija. Ši stabilizacija įvyksta greičiau nei praeitame modelyje. Diskriminatoriaus praradimai iteracijų pradžioje taip pat didesni, tačiau greitai stabilizuojasi ties 0.5-1 reikšme, panašiai kaip ir *StyleGAN2-ADA* modelis. Iš grafiko matoma, kad ši modelio versija geba tolygiau apmokyti generatorių ir diskriminatorių, kadangi stabilizacija įvyksta anksčiau ir tolygiau – nėra didelio reikšmių svyravimo iteracijų pradžioje.



19 pav. *StyleGAN3* generatoriaus ir diskriminatoriaus praradimai apmokymo metu (*Blazer* klasė)

Apmokius modelius, generuojami paveikslėliai duomenų rinkinio papildymui. Žemiau pateikti *Blazer* klasės generuoti paveikslėlių pavyzdžiai (žr. 20 pav.). Generuotuose paveikslėliuose vizualiai matosi klasės savybės, iš kurių galima spręsti kad tai yra *Blazer* klasės paveikslėlis, lyginant su *StyleGAN2-ADA* generuotais paveikslėliais, jie atrodo gana panašiai, taip pat aiškiai matosi, kad tai nėra tikros drabužių nuotraukos, tačiau paveikslėliuose pastebima mažiau nerealistiškai atrodančių artefaktų.



20 pav. *StyleGAN3* *Blazer* klasės sugeneruotų paveikslėlių pavyzdys

2.3. Hierarchinių modelių architektūra

Bus apmokomi trijų variacijų hierarchiniai modeliai. Modeliai parinkti grupuojant klases pagal klasių vizualinę informaciją.

2.3.1. Trijų lygių hierarchinis modelis

Trijų lygių hierarchinis modelis (žr. **3 lentelė**) projektuojamas atsižvelgiant į vizualias klasių paveikslėlių savybes tokias kaip: kūno dalis ant kurios rūbas nešiojamas, rankovės ar bendras ilgis. Klasės pirmiausiai padalijamos į tris kategorijas: apatinės dalies drabužius (*lower-body*), viršutinės dalies drabužius (*upper-body*) ir viso kūno ilgio drabužius (*full-body*). Šios klasės išskaidytos pagal vaizdines savybes ir kategorijos atitikmenį. Apatinės dalies drabužiams priskirti šortai (*shorts*), kelnės (*pants*) ir sijonai (*skirt*). Viršutinės dalies drabužiams priskirti marškinėliai (*tops*), megztiniai (*sweaters*) ir lauko drabužiai (*outwear*). Pilno ilgio drabužiams priskirti vienos dalies drabužiai (*one-piece*) ir lauko viso ilgio drabužiai (*outer*). Šios klasės atitinkamai išskaidytos į smulkesnes atitinkamas kategorijas – šortams (*shorts*) priskirti visi šortų tipai: *shorts*, *cutoffs*, *sweatshorts*. Kelnėms priskirti visi kelnų tipai: *jeans*, *leggings*, *joggers*, *sweatpants*. Sijonams nepriskirta daugiau klasių, kadangi jie neturi atitinkamų susijusių klasių. Marškinėliams priskirtos *tee*, *blouse*, *tank*, *top* klasės. Megztiniams priskirtos *cardigan*, *sweater*, *hoodie* klasės. Lauko drabužiams priskirtos *jacket* ir *blazer* klasės. Viso ilgio drabužiams priskirtos *dress*, *romper*, *jumpsuit* klasės. Viso ilgio lauko drabužiams priskirtos *kimono* ir *coat* klasės.

3 lentelė. Trijų hierarchinių lygių mazgų architektūra

Pirmojo lygio mazgo klasė	Antrojo lygio mazgo klasė	Trečiojo lygio mazgo klasės
<i>lower-body</i>	<i>shorts</i>	<i>shorts, cutoffs, sweatshorts</i>
	<i>pants</i>	<i>jeans, leggings, joggers, sweatpants</i>
	<i>skirt</i>	
<i>upper-body</i>	<i>tops</i>	<i>tee, blouse, tank, top</i>
	<i>sweaters</i>	<i>cardigan, sweater, hoodie</i>
	<i>outwear</i>	<i>jacket, blazer</i>
<i>full-body</i>	<i>one-piece</i>	<i>dress, romper, jumpsuit</i>
	<i>outer</i>	<i>kimono, coat</i>

2.3.2. Dviejų lygių hierarchinis modelis

Dviejų lygių hierarchinis modelis (žr. **4 lentelė**) projektuojamas atsižvelgiant į vizualias klasių paveikslėlių savybes tokias kaip: kūno dalis ant kurios rūbas nešiojamas, bendras ilgis, atmetant rankovės ilgio savybę ir nebeišskiriant sijono kaip atskiro mazgo. Klasės padalijamos į tris kategorijas: apatinės dalies drabužius (*lower-body*), viršutinės dalies drabužius (*upper-body*) ir viso kūno ilgio drabužius (*full-body*). Šios klasės išskaidytos pagal vaizdines savybes ir kategorijos atitikmenį. Apatinės dalies drabužiams priskirtos visos šortų variacijos kartu su sijonais ir kelnų variacijomis: *shorts*, *cutoffs*, *sweatshorts*, *jeans*, *leggings*, *joggers*, *sweatpants*, *skirt*. Viršutinės dalies drabužiams priskirti visi drabužiai kurie užsivelkami ant viršutinės kūno dalies ar turi rankoves: *coat*, *tee*, *blouse*, *tank*, *top*, *cardigan*, *sweater*, *hoodie*, *jacket*, *blazer*. Pilno ilgio drabužiams priskirti vienos dalies drabužiai, kurie vizualiai primena sukneles ar yra ištisiniai: *dress*, *romper*, *jumpsuit*, *kimono*.

4 lentelė. Dviejų hierarchinių lygių mazgų architektūra

Pirmojo lygio mazgo klasė	Antrojo lygio mazgo klasės
<i>lower-body</i>	<i>shorts, cutoffs, sweatshorts, jeans, leggings, joggers, sweatpants, skirt</i>
<i>upper-body</i>	<i>coat, tee, blouse, tank, top, cardigan, sweater hoodie, jacket, blazer</i>
<i>full-body</i>	<i>dress, romper, jumpsuit, kimono</i>

2.3.3. Trijų nepilnų lygių hierarchinis modelis

Trijų nepilnų lygių hierarchinis modelis (žr. **5 lentelė**) projektuojamas atsižvelgiant į vizualias klasių paveikslėlių savybes tokias kaip: kūno dalis ant kurios rūbas nešiojamas, bendras ilgis. Modelis yra išplėsta dviejų hierarchinių lygių variacija, kurioje apatinės dalies drabužiai papildomai išskaidyti į kelnių ir šortų kartu su sijonais kategorijas.

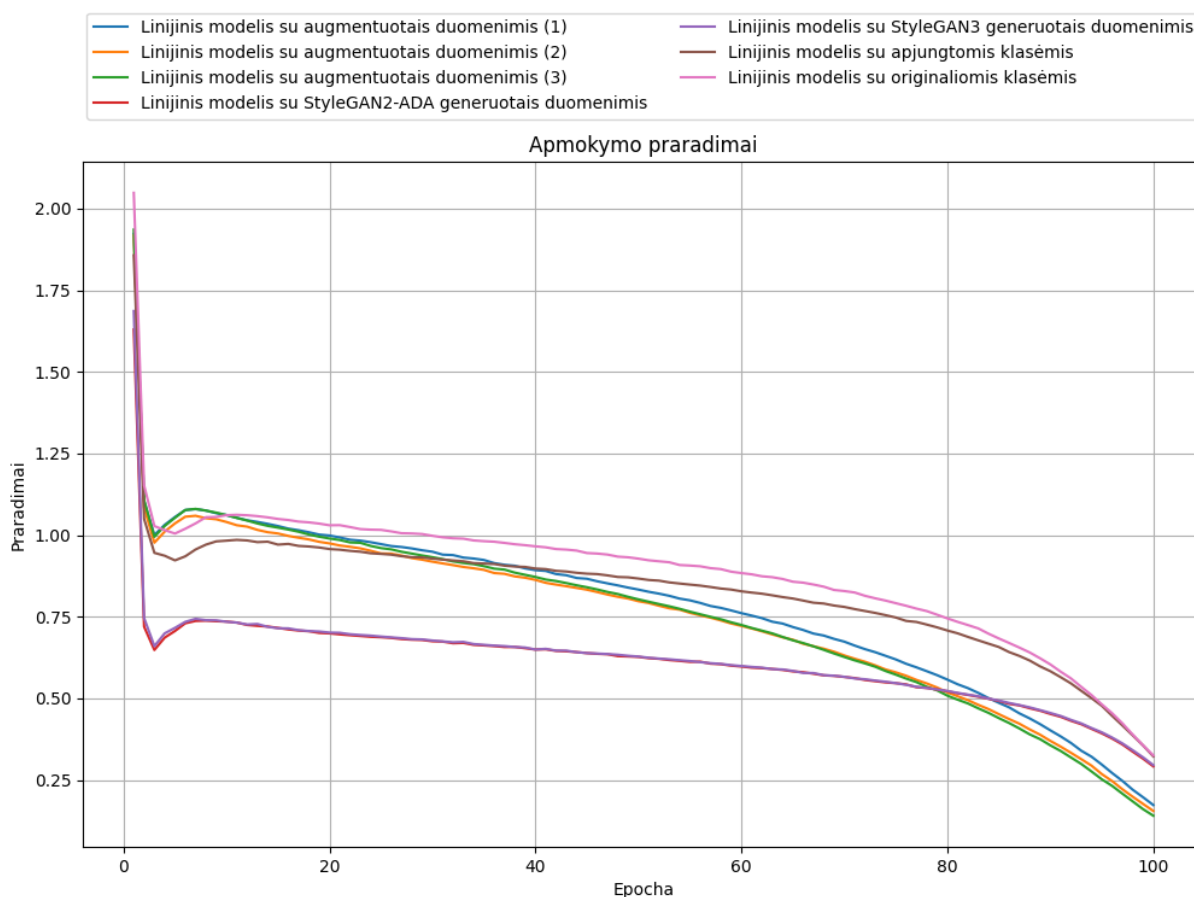
5 lentelė. Trijų nepilnų hierarchinių lygių mazgų architektūra

Pirmojo lygio mazgo klasė	Antrojo lygio mazgo klasė	Trečiojo lygio mazgo klasės
<i>lower-body</i>	<i>other</i>	<i>shorts, cutoffs, sweatshorts, skirt</i>
	<i>pants</i>	<i>jeans, leggings, joggers, sweatpants</i>
<i>upper-body</i>	<i>coat, tee, blouse, tank, top, cardigan, sweater, hoodie, jacket, blazer</i>	
<i>full-body</i>	<i>dress, romper, jumpsuit, kimono</i>	

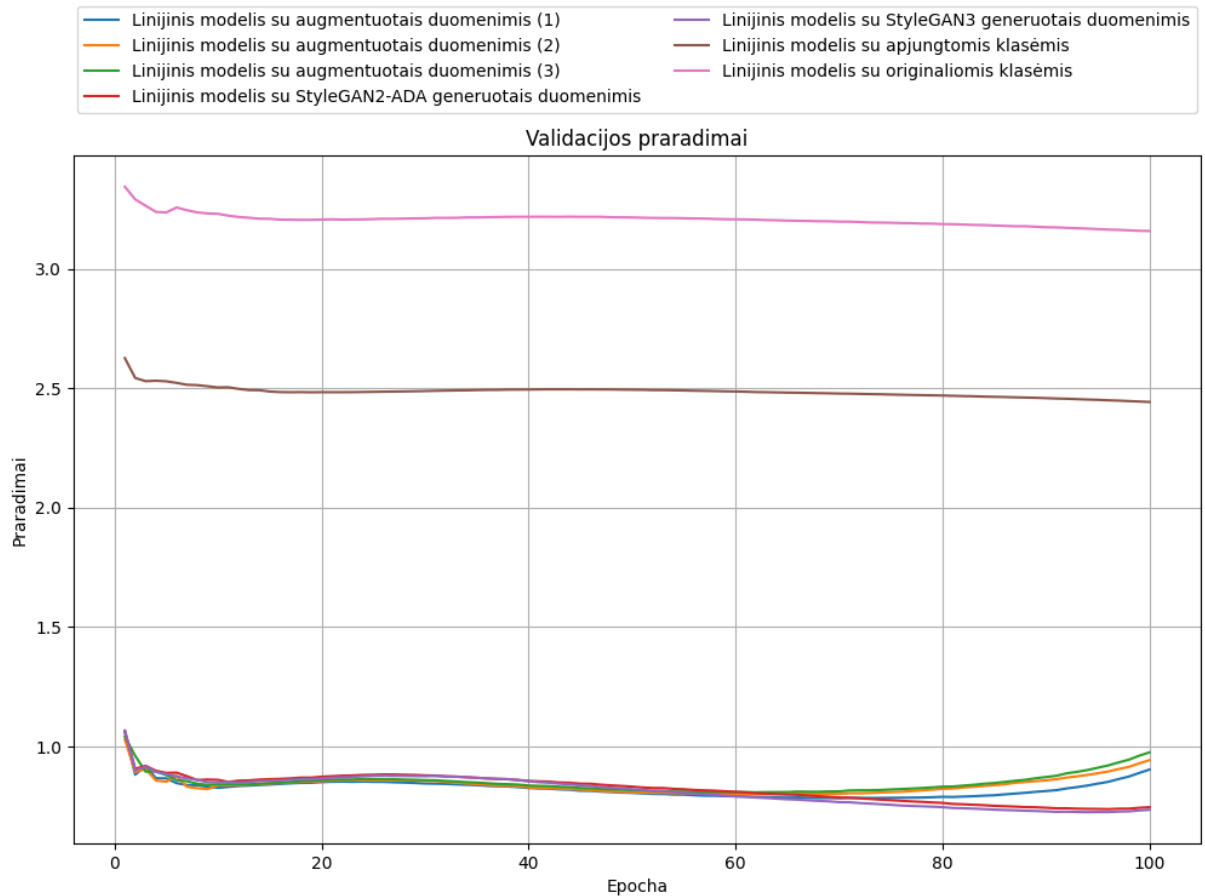
3. Eksperimentai atlikti mados vaizdų atpažinimo sprendimui

Eksperimentai atliekami panaudojant *YOLOv11-cls* klasifikacinio modelio architektūrą. Atliekami bandymai su įvairiomis duomenų rinkinio modifikacijomis, pritaikant jas modelių variacijoms. Modeliai apmokomi 100 epochų – su šiuo kiekiu modeliai apsimokė geriausiai. Naudojamos geriausius rezultatus pasiekusios per nurodytą kiekį epochų modelių versijos.

Žemiau pateikiami apmokytų linijinių modelių apmokymo praradimų (žr. **21 pav.**) ir validacijos praradimų (žr. **22 pav.**) grafikai. Apmokymo praradimų grafike modelių praradimai pirmųjų epochų metu staigiai krenta, tačiau didžiausias skirtumas matomas tarp GAN metodais išplėstų duomenų rinkinių modelių ir kitų modelių. GAN generuotų duomenų modeliai pasiekė žymiai mažesnius apmokymo praradimus. Nors ties 100 epocha matoma, kad augmenteduotais duomenimis išplėstų modelių apmokymo praradimai yra žemiausi, tačiau iš validacijos praradimų grafiko matoma, kad šie modeliai paskutinėmis epochomis pradėjo persimokyti – jų validacijos praradimai pradėjo kilti. Žymus skirtumas matomas ir validacijos praradimų grafike, kur nepraplėstų duomenų modeliai turi didžiausius praradimus lyginant su kitais modeliais. Galima daryti išvadą, kad mažų klasių duomenų išplėtimas, duoda didelę naudą modelio apmokymo rezultatams – modelis išmoksta geriau generalizuoti klases.



21 pav. Apmokymo praradimai



22 pav. Validacijos praradimai

3.1. Linijinio *YOLOv11-cls* modelio eksperimentai su nepraplėstu duomenų rinkiniu

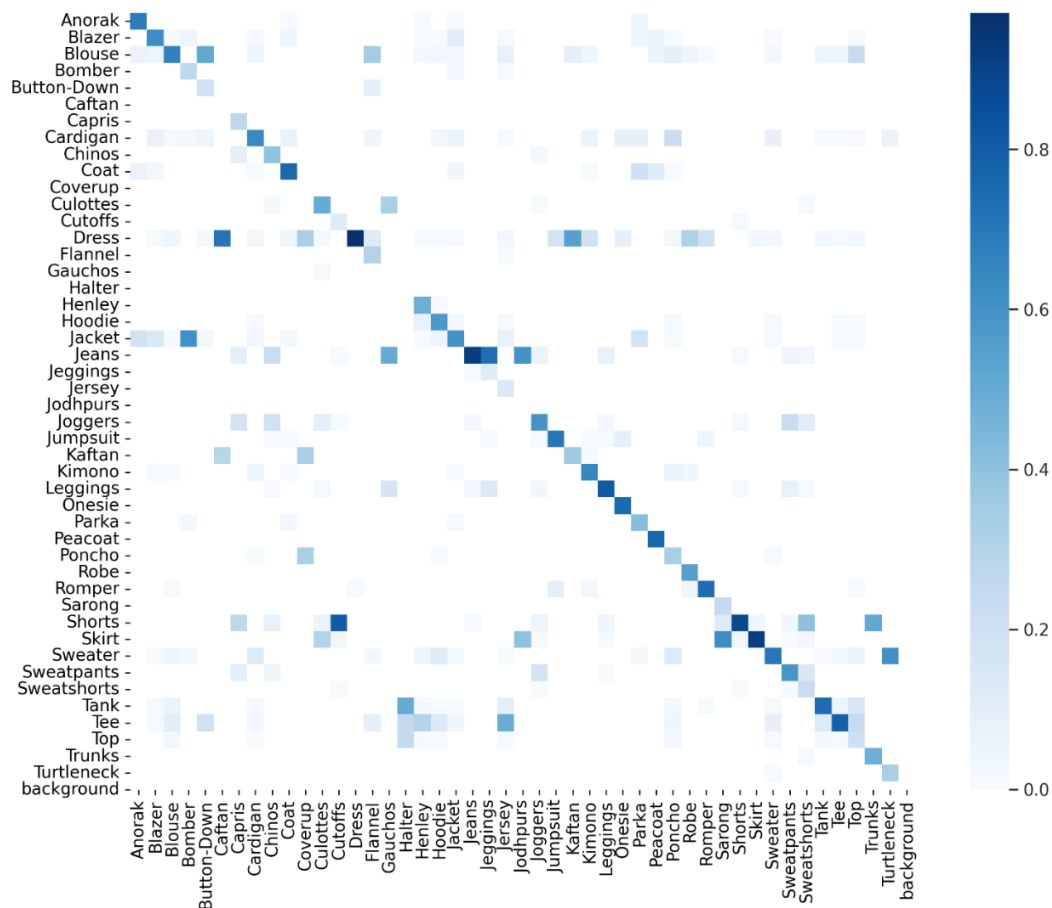
Šiems eksperimentams paruošiami du duomenų rinkiniai:

1. Pilnas duomenų rinkinys.
2. Duomenų rinkinys, kurio mažiausios duomenų klasės apjungtos su jas atitinkančiomis didesnėmis klasėmis.

3.1.1. Linijinio modelio su pilnu duomenų rinkiniu rezultatai:

Pirmaisiais bandymais apmokytas linijinis modelis, panaudojant viso duomenų rinkinio klases. Šiuo bandymu buvo pagerintas esamas geriausias *DeepFashion* duomenų rinkinio klasifikavimo tikslumas – pasiektas 76.15% *top-1* tikslumas, 92.59% *top-3* tikslumas ir 96.60% *top-5* tikslumas (žr. **6 lentelė**). Pasiiektas 76.15% tikslumas, 60.96% preciziškumas, 47.91% jautrumas ir 49.01% *F1* (žr. **21 lentelė**).

Susimaišymo matricos lentelėje (žr. **23 pav.**) matoma, kad kai kurios klasės tarpusavyje stipriai koreliuoja. Šios klasės tarpusavyje yra labai panašios, todėl modelis sunkiai prognozuoja kuri klasė yra tikroji.



23 pav. Linijinio YOLOv11 modelio, apmokyto su visomis duomenų rinkinio klasėmis, susimaišymo matrica

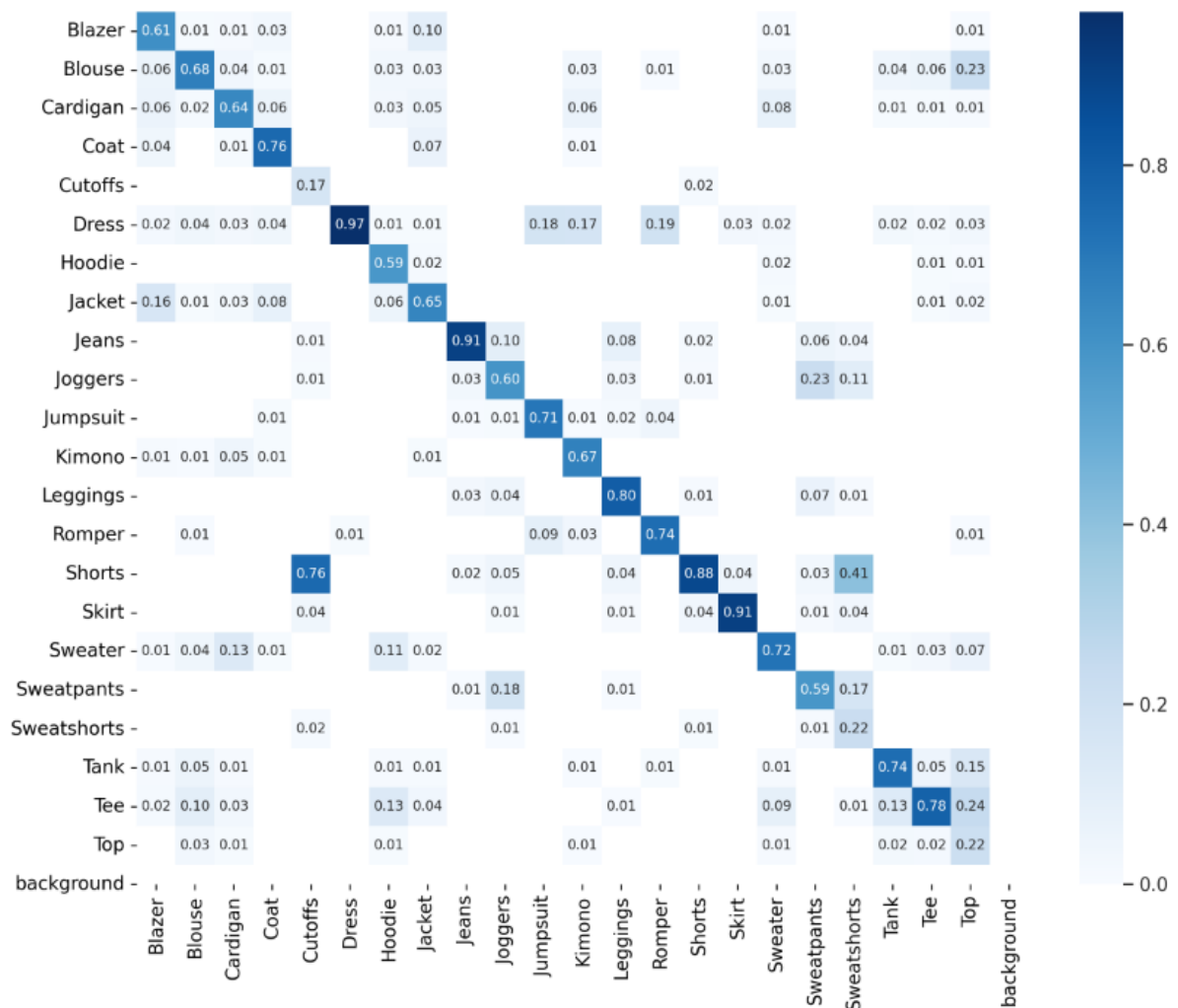
6 lentelė. Linijinio modelio, klasifikuojančio neapjungtų klasių duomenų rinkinį, rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Linijinis modelis su originaliomis klasėmis	76.15	92.59	96.60

3.1.2. Linijinio modelio su apjungtu duomenų rinkiniu rezultatai:

Didesnio tikslumo pasiekimui, nuspręsta išmesti labai mažo dydžio klases arba apjungti jas į bendrą duomenų rinkinį. Atlikus šio duomenų rinkinio paruošimą apmokytas linijinis modelis, kuris pagerino prieš modelio, kuriame naudojamos visos duomenų rinkinio klasės tikslumą (žr. 7 lentelė). Klasės, kurios buvo apjungtos, tarpusavyje stipriai koreliavo, todėl jas pašalinus ar apjungus buvo galima pasiekti didesnę tikslumą. Pasiektas 77.33% top-1 tikslumas, 93.36% top-3 tikslumas ir 97.26% top-5 tikslumas (žr. 7 lentelė). Pasiektas 77.33% tikslumas, 68.98% preciziškumas, 66.33% jautrumas ir 66.90% F1 (žr. 21 lentelė).

Susimaišymo matricoje (žr. 24 pav.) pastebimi geresni rezultatai – sumažėjo skirtingų klasių, kurios tarpusavyje koreliuoja kiekis.



24 pav. Linijinio YOLOv11 modelio, apmokyto su apjungtomis duomenų rinkinio klasėmis, susimaišymo matrica

7 lentelė. Linijinio modelio, klasifikuojančio apjungtų klasių duomenų rinkinį, rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Linijinis modelis su apjungtomis klasėmis	77.33	93.36	97.26

3.2. Hierarchinio YOLOv11-cls modelio eksperimentai

Aukštesnio tikslumo pasiekimui, pasitelkta hierarchinio modelio architektūra, kurios kiekvienas mazgas apmokytas nuspėti mažesnę kiekį klasių. Aukštesnio lygio mazgų išvestis nulemia kuris žemesnio lygio mazgas prognozuos tolimesnę reikšmę. Šis sprendimo būdas leidžia apmokyti kiekvieną mazgą konkretesnei užduočiai – kiekvienas modelis apmokomas nuspėti mažesnę kiekį klasių.

Šiems eksperimentams paruošiami trys duomenų rinkiniai:

1. Trijų lygių hierarchijos duomenų rinkinys
2. Trijų nepilnų lygių hierarchijos duomenų rinkinys
3. Dviejų lygių hierarchijos duomenų rinkinys

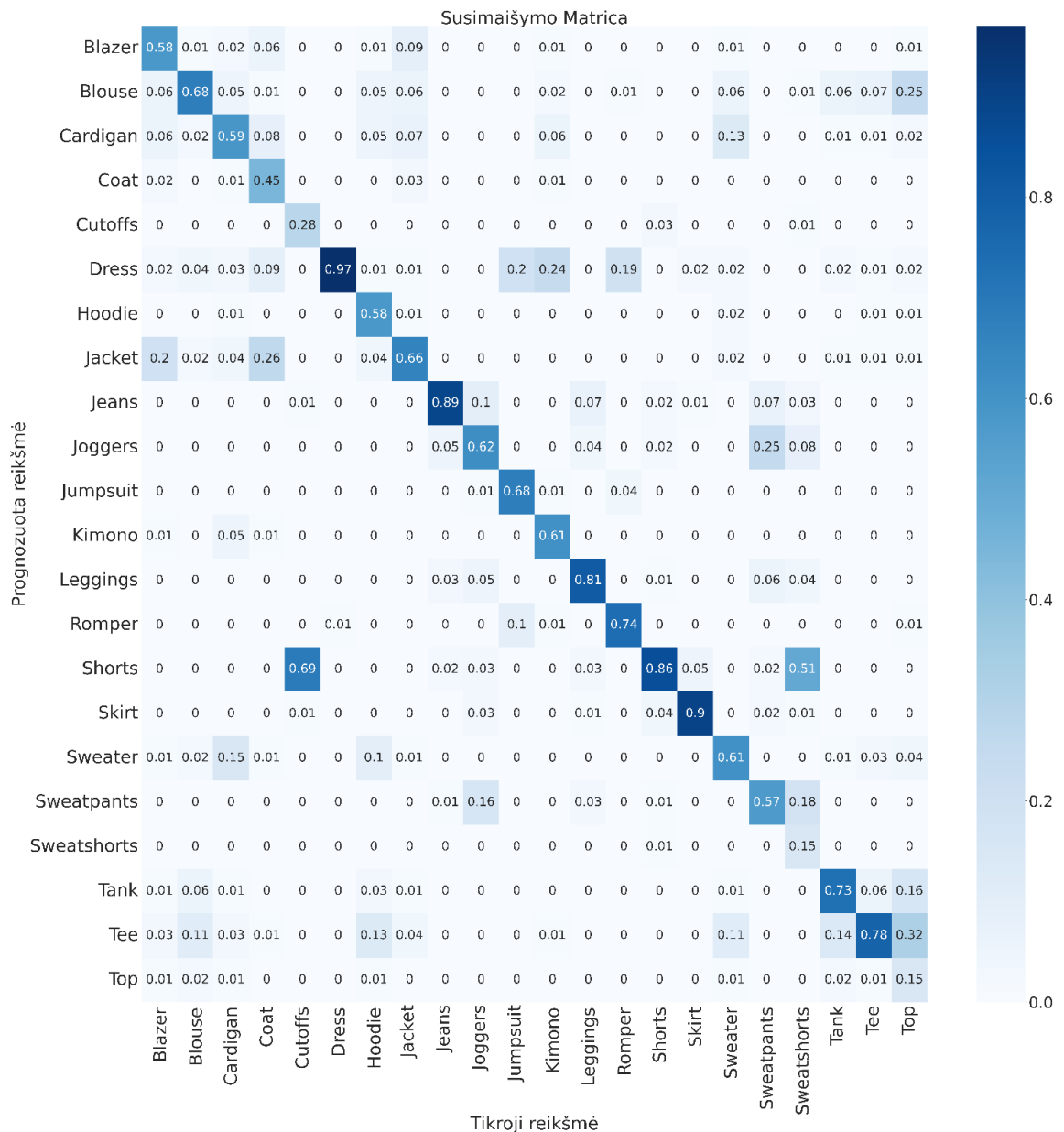
3.2.1. Trijų lygių hierarchinio modelio rezultatai

Apmokyto hierarchinio modelio bendras tikslumas: bendras lapinių kategorijų *top-1* – 76.08%, *top-3* – 88.78%, *top-5* – 89.88% (žr. **9 lentelė**). Apmokyto trijų lygių hierarchinio modelio mazgai pasiekė aukštą tikslumą (žr. **8 lentelė**) didžioji dalis *top-3* ir *top-5* reikšmių lygios 1, tačiau tai yra dėl to, kadangi kiekvieno mazgo išvestis yra 3-4 klasės. Nors atskirų mazgų tikslumas yra aukštas, tačiau juos apjungus į bendrą hierarchinį modelį, tikslumas sumažėja, kadangi netikslumai aukštesnio lygio klasifikavime nulemia žemesnių mazgų netinkamą pasirinkimą. Pasiektas bendras 76.08% tikslumas, 67.45% preciziškumas, 63.49% jautrumas ir 64.72% *F1* (žr. **21 lentelė**).

Susimaišymo matricoje matomas susimaišymo padidėjimas (žr. **25 pav.**). Padidėjo *Blazer, Cardigan, Coat, Hoodie, Jeans, Jumpsuit, Kimono, Shorts, Skirt, Sweater, Sweatpants, Sweatshorts, Tank, Top* klasių susimaišymas. Sumažėjo *Cutoffs, Jacket, Joggers, Leggings* klasių susimaišymas, tarp kurių trys klasės priklauso mažiau nei 10000 paveikslėlių turinčioms klasėms. Nepasikeitė *Blouse, Dress, Romper, Tee* klasių susimaišymas. Bendras susimaišymas padidėjo didžiajai daliai mažesnę kiekį paveikslėlių turinčioms klasėms.

8 lentelė. Trijų lygių hierarchinio modelio mazgų rezultatai

Hierarchinio modelio mazgas	Hierarchinio modelio mazgo lygis	<i>top-1 (%)</i>	<i>top-3 (%)</i>	<i>top-5 (%)</i>
Pirmojo lygio mazgas	1	97.25	1.00	1.00
Antrojo lygio apatinės dalies rūbų mazgas	2	92.70	1.00	1.00
Antrojo lygio viršutinės dalies rūbų mazgas		88.22	1.00	1.00
Antrojo lygio viso kūno rūbų mazgas		98.14	1.00	1.00
Trečiojo lygio vienos dalies rūbų mazgas	3	94.48	1.00	1.00
Trečiojo lygio viršutinių rūbų mazgas		95.44	1.00	1.00
Trečiojo lygio palaidinių mazgas		71.50	96.20	1.00
Trečiojo lygio megztinių mazgas		76.04	1.00	1.00
Trečiojo lygio lauko drabužių mazgas		79.22	1.00	1.00
Trečiojo lygio šortų mazgas		87.74	1.00	1.00
Trečiojo lygio kelnių mazgas		79.14	97.82	1.00



25 pav. Trijų lygių hierarchinio YOLOv11 modelio susimaišymo matrica

9 lentelė. Trijų lygių hierarchinio modelio rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Trijų lygių hierarchinis modelis	76.08	88.78	89.88

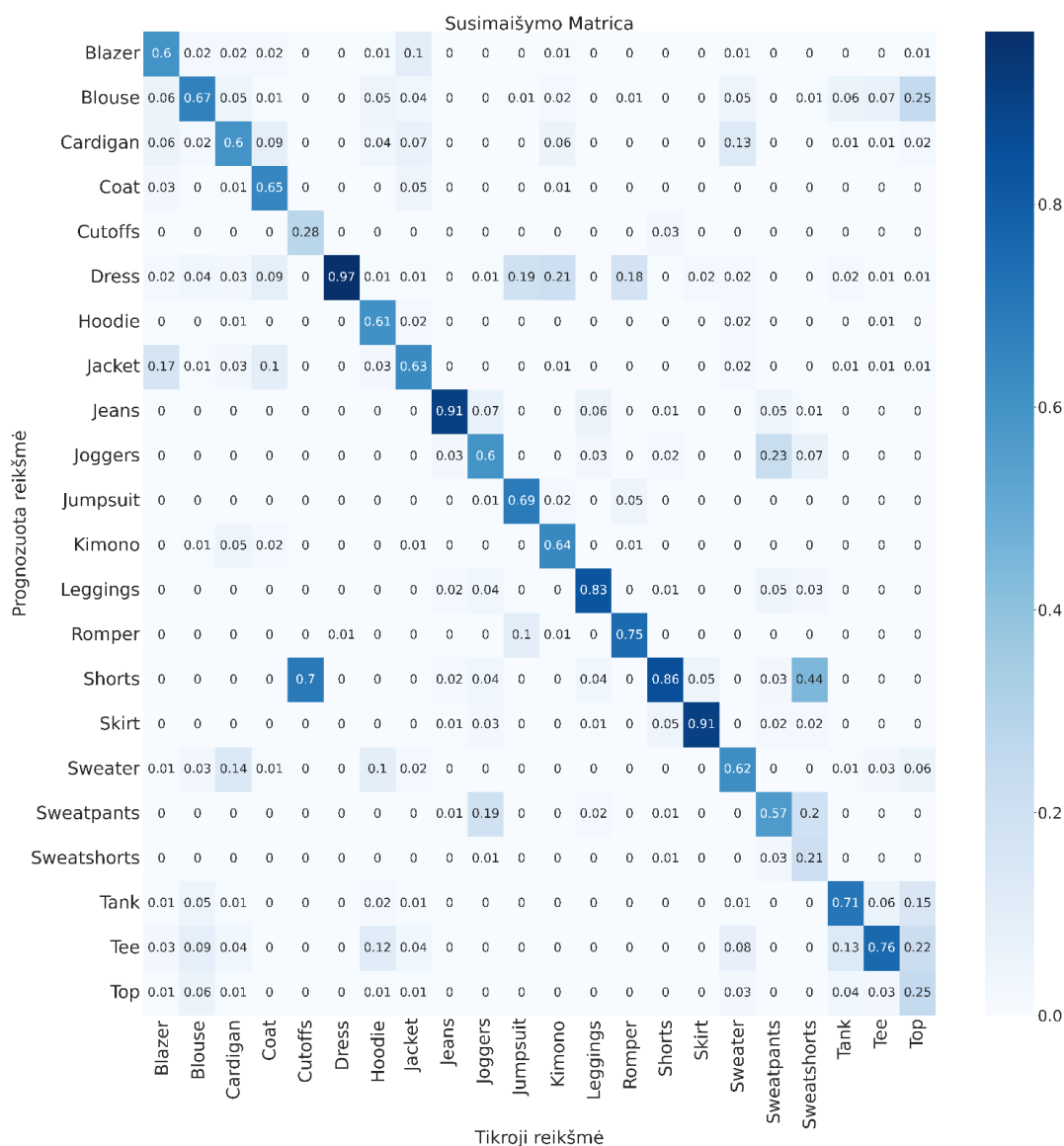
3.2.2. Dviejų lygių hierarchinio modelio rezultatai

Apmokyto hierarchinio modelio bendras tikslumas: bendras lapinių kategorijų *top-1* – 76.53%, *top-3* – 91.78%, *top-5* - 95.40% (žr. 11 lentelė). Apmokyto dviejų lygių hierarchinio modelio mazgai pasiekė aukštą tikslumą (žr. 10 lentelė). Kai kurios *top-3* ir *top-5* reikšmės yra lygios 1, tačiau tai yra dėl to, kadangi šių mazgų išvestys tame lygyje yra 3-4 klasės. Lyginant su trijų lygių hierarchiniu modeliu, šio modelio atskirų mazgų tikslumas pirmame lygyje yra didesnis, tačiau antrojo lygio mazgų tikslumas, lyginant su trijų lygių hierarchinio modelio antru ir trečiu lygiais yra žymiai prastesnis. Pasiektas bendras 76.53% tikslumas, 67.05% preciziškumas, 65.18% jautrumas ir 65.76% *F1* (žr. 21 lentelė).

Susimaišymo matricoje matomas susimaišymo sumažėjimas (žr. 26 pav.). Padidėjo *Blazer*, *Blouse*, *Jacket*, *Joggers*, *Tank*, *Tee* klasių susimaišymas. Sumažėjo *Cardigan*, *Coat*, *Hoodie*, *Jeans*, *Jumpsuit*, *Kimono*, *Leggings*, *Romper*, *Skirt*, *Sweater* klasių susimaišymas. Nepasikeitė *Cutoffs*, *Dress*, *Shorts*, *Sweatpants*, *Sweatshorts*, *Top* klasių susimaišymas. Bendras susimaišymas sumažėjo didžiajai daliai mažesnę kiekį paveikslėlių turinčioms klasėms, padidėjo didesnę kiekį paveikslėlių turinčioms klasėms.

10 lentelė. Dviejų lygių hierarchinio modelio mazgų rezultatai

Hierarchinio modelio mazgas	Hierarchinio modelio mazgo lygis	top-1 (%)	top-3 (%)	top-5 (%)
Pirmojo lygio mazgas	1	97.54	1.00	1.00
Antrojo lygio apatinės dalies rūbų mazgas	2	81.95	96.45	98.99
Antrojo lygio viršutinės dalies rūbų mazgas		68.05	90.22	96.38
Antrojo lygio viso kūno rūbų mazgas		93.72	99.56	1.00



26 pav. Dviejų lygių hierarchinio YOLOv11 modelio susimaišymo matrica

11 lentelė. Dviejų lygių hierarchinio modelio rezultatai

Modelis	<i>top-1 (%)</i>	<i>top-3 (%)</i>	<i>top-5 (%)</i>
Dviejų lygių hierarchinis modelis	76.53	91.78	95.40

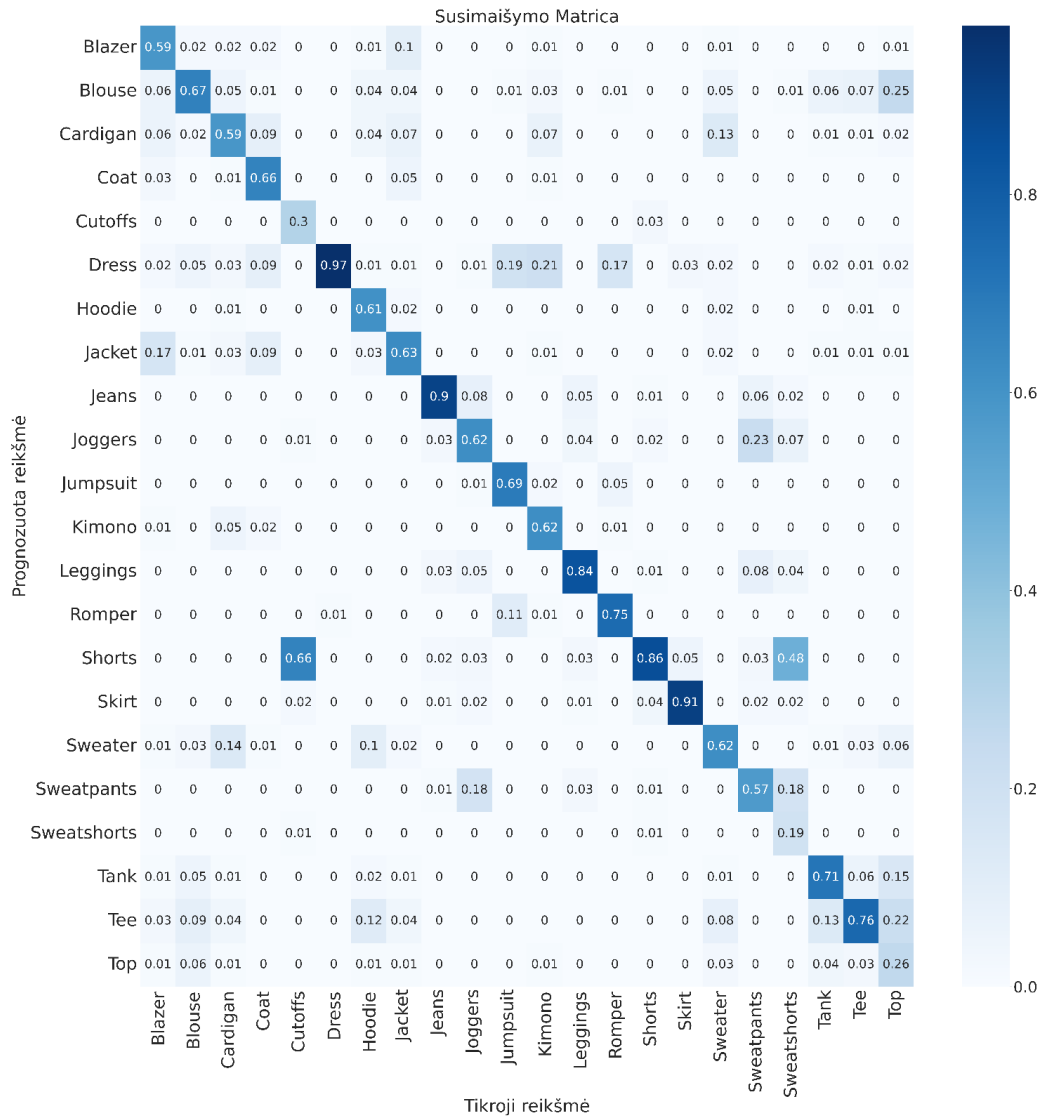
3.2.3. Trijų nepilnų lygių hierarchinio modelio rezultatai

Apmokyto hierarchinio modelio bendras tikslumas - Bendras lapinių kategorijų *top-1* – 76.53%, *top-3* – 91.39%, *top-5* - 94.73% (žr. **13 lentelė**). Apmokyto trijų nepilnų lygių hierarchinio modelio mazgai (žr. **12 lentelė**) pasiekė dar žemesnį tikslumą, nei trijų ar dviejų lygių hierarchiniai modeliai vertinant *top-3* ir *top-5* metrikas. Nors *top-1* metrika ir sutampa su dviejų lygių hierarchiniu modeliu (žr. **11 lentelė**). Kai kurios *top-3* ir *top-5* reikšmės yra lygios 1, tačiau tai yra dėl to, kadangi šių mazgų išvestys tame lygyje yra 2-4 klasės. Lyginant su trijų lygių hierarchiniu modeliu, šio modelio atskirų mazgų tikslumas pirmame lygyje yra didesnis, tačiau antrojo lygio mazgų tikslumas, lyginant su trijų lygių hierarchinio modelio antru ir trečiu lygiais yra žymiai prastesnis. Pasiektas bendras 76.53% tikslumas, 67.50% preciziškumas, 65.10% jautrumas ir 65.72% *F1* (žr. **21 lentelė**).

Susimaišymo matricoje matomas susimaišymo padidėjimas (žr **27 pav.**). Lyginant su linijiniu apjungtų klasių modeliu, padidėjo *Blazer, Cardigan, Jeans, Kimono, Sweatshorts* klasių susimaišymas. Sumažėjo *Coat, Cutoffs, Joggers, Leggings, Top* klasių susimaišymas. Nepasikeitė *Blouse, Dress, Hoodie, Jacket, Jumpsuit, Romper, Shorts, Skirt, Sweater, Sweatpants, Tank, Tee* klasių susimaišymas. Bendras susimaišymas sumažėjo ir padidėjo mažą kiekį paveikslėlių turinčioms klasėms. Didesnį kiekį paveikslėlių turinčių klasių susimaišymas beveik nepakitė.

12 lentelė. Trijų nepilnų lygių hierarchinio modelio mazgų rezultatai

Hierarchinio modelio mazgas	Hierarchinio modelio mazgo lygis	<i>top-1 (%)</i>	<i>top-3 (%)</i>	<i>top-5 (%)</i>
Pirmojo lygio mazgas	1	97.50	1	1
Antrojo lygio apatinės dalies rūbų mazgas	2	95.61	1	1
Antrojo lygio viršutinės dalies rūbų mazgas		68.05	90.22	96.38
Antrojo lygio viso kūno rūbų mazgas		93.72	99.56	1
Trečiojo lygio apatinės dalies rūbų virš kelių ilgio mazgas	3	87.01	99.37	1
Trečiojo lygio apatinės dalies rūbų žemiau kelių ilgio mazgas		95.44	97.96	1



27 pav. Trijų nepilnų lygių hierarchinio YOLOv11 modelio susimaišymo matrica

13 lentelė. Trijų nepilnų lygių hierarchinio modelio rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Trijų nepilnų lygių hierarchinis modelis	76.53	91.39	94.73

3.3. Eksperimentai su GAN metodais išplestu apmokymo duomenų rinkiniu

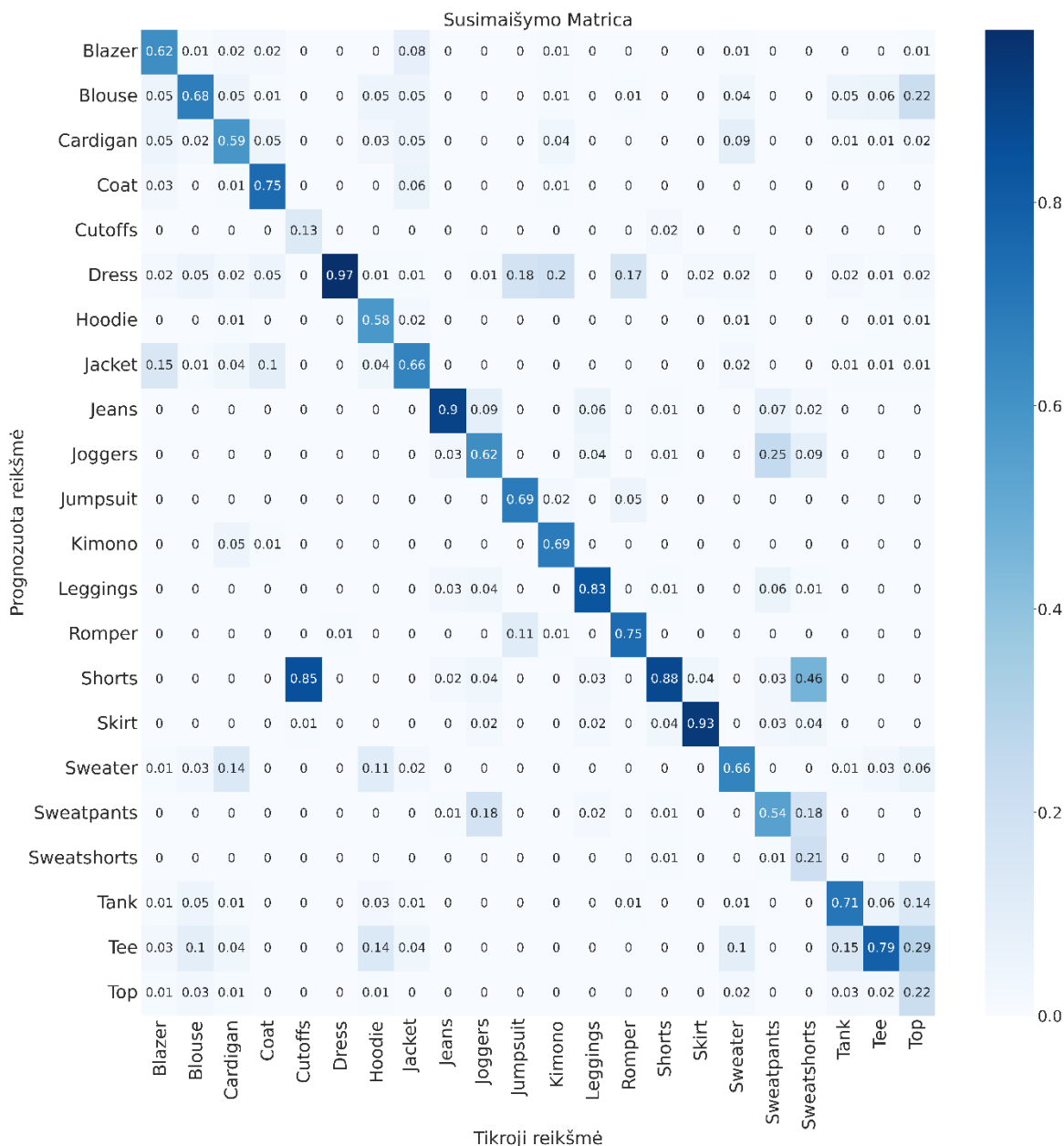
Eksperimentai atliekami panaudojant YOLOv11-cls klasifikacinio modelio architektūrą. Bandymai atlikti su GAN metodais išplėstu duomenų rinkiniu.

3.3.1. YOLOv11-cls modelio eksperimentai su StyleGAN2-ADA metodo papildytu duomenų rinkiniu

Apmokyto modelio bendras tikslumas - top-1 – 77.30%, top-3 – 93.44%, top-5 - 97.34% (žr. 14 lentelė). Šis modelis pagerino linijinio modelio su apjungtomis duomenų klasėmis top-3 ir top-5 metrikas. Tačiau nors ir lyginamosios metrikos pagerėjo, bet kitos tikslumo metrikos, tokios kaip tikslumas, preciziškumas, jautrumas ir F1 šiam modeliui yra prastesnės. Šie rezultatai rodo, kad apmokius modelį su generuotais duomenimis išplėstu duomenų rinkiniu, modelis išmoko geriau

generalizuoti klases, tačiau nuo to suprastėjo modelio patikimumas prognozuojant tikrąją klasę. Pasiektas 77.30% tikslumas, 68.95% preciziškumas, 65.43% jautrumas ir 66.15% *F1* (žr. **21 lentelė**).

Susimaišymo matricoje matomas susimaišymo pakitimas (žr. **28 pav.**). Lyginant su linijiniu apjungtų klasių modeliu, padidėjo *Cardigan*, *Coat*, *Cutoffs*, *Hoodie*, *Jeans*, *Joggers*, *Jumpsuit*, *Sweater*, *Sweatpants*, *Sweatshorts*, *Tank* klasių susimaišymas. Sumažėjo *Blazer*, *Jacket*, *Kimono*, *Leggings*, *Romper*, *Skirt*, *Tee* klasių susimaišymas. Nepasikeitė *Blouse*, *Dress*, *Shorts*, *Top* klasių susimaišymas. Didesnį kiekį paveikslėlių turinčių klasių susimaišymas kito mažai, kitų klasių susimaišymas kito į abi puses.



28 pav. Linijinio modelio su *StyleGAN2-ADA* generuotais duomenimis susimaišymo matrica

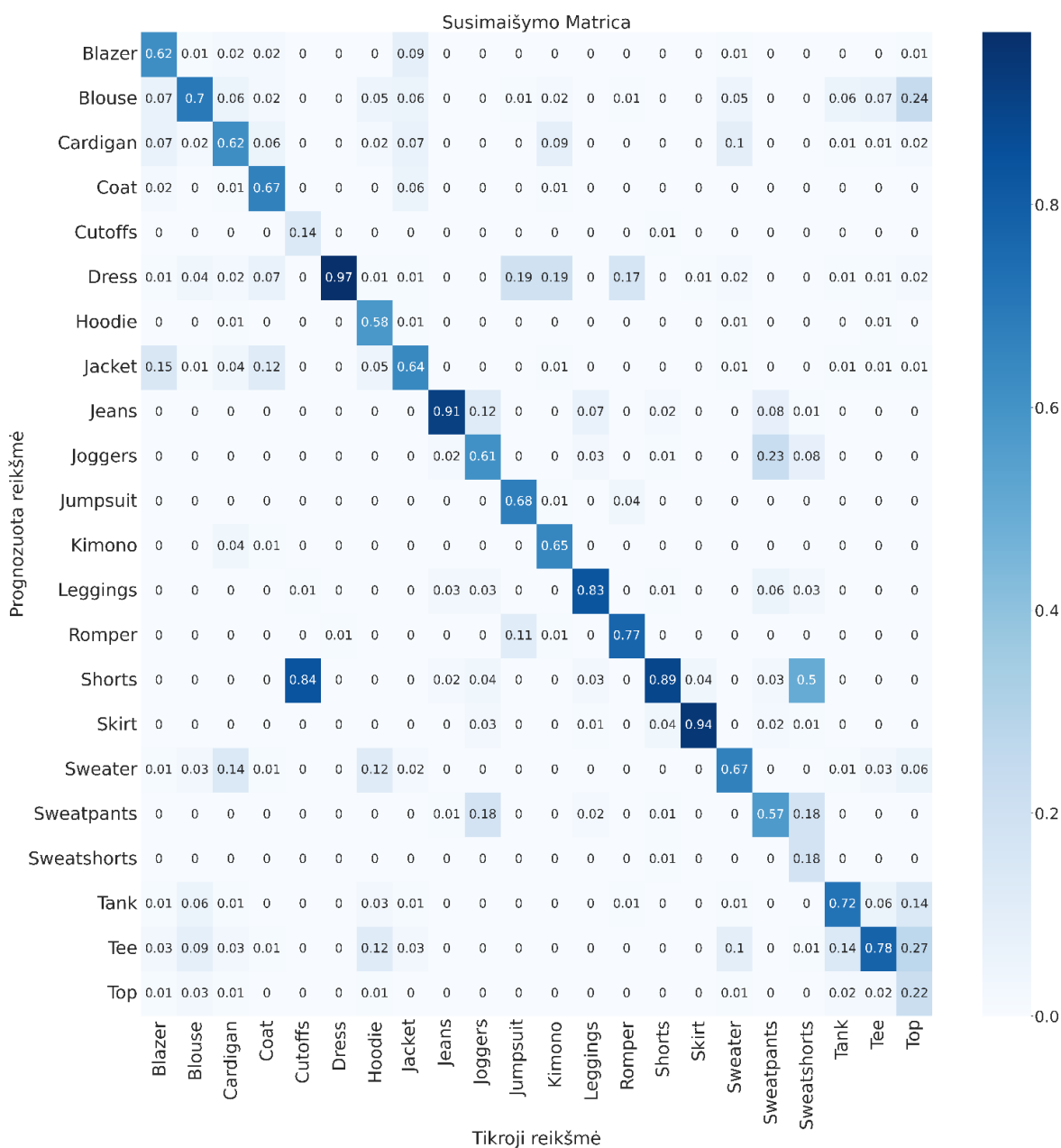
14 lentelė. Linijinio modelio su *StyleGAN2-ADA* generuotais duomenimis rezultatai

Modelis	<i>top-1 (%)</i>	<i>top-3 (%)</i>	<i>top-5 (%)</i>
Linijinis modelis su <i>StyleGAN2-ADA</i> generuotais duomenimis	77.30	93.44	97.34

3.3.2. YOLOv11-cls modelio eksperimentai su StyleGAN3 metodo papildytu duomenų rinkiniui

Apmokytą modelio bendras tikslumas - *top-1* – 77.53%, *top-3* – 93.48%, *top5* - 97.40% (žr. **15 lentelė**). Šis modelis pagerino tiek linijinio modelio su apjungtomis klasėmis, tiek su *StyleGAN2-ADA* papildyto modelio *top-k* metrikas. Modelis taip pat pagerino visų kitų modelių svorines tikslumo, preciziškumo, jautrumo ir *F1* metrikas, makro tikslumo ir preciziškumo metrikas tačiau makro jautrumas ir *F1* aukščiausias išliko linijinio modelio su apjungtomis klasėmis. Pasiektas 87.49% tikslumas, 76.71% preciziškumas, 77.53% jautrumas ir 76.69% *F1* (žr. **21 lentelė**).

Susimaišymo matricoje matomas susimaišymo pakitimas (žr **29 pav.**). Lyginant su linijiniu apjungtų klasių modeliu, padidėjo *Cardigan*, *Coat*, *Cutoffs*, *Hoodie*, *Jacket*, *Jumpsuit*, *Kimono*, *Sweater*, *Sweatpants*, *Sweatshorts*, *Tank* klasių susimaišymas. Sumažėjo *Blazer*, *Blouse*, *Joggers*, *Leggings*, *Romper*, *Shorts*, *Skirt* klasių susimaišymas. Nepasikeitė *Dress*, *Jeans*, *Tee*, *Top* klasių susimaišymas.



29 pav. Linijinio modelio su *StyleGAN3* generuotais duomenimis susimaišymo matrica

15 lentelė. Linijinio modelio su *StyleGAN3* generuotais duomenimis rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Linijinis modelis su <i>StyleGAN3</i> generuotais duomenimis	77.53	93.48	97.40

3.4. *YOLOv11-cls* modelio eksperimentai su augmentacijos metodais papildytu duomenų rinkiniui

Atlikti bandymai su trimis augmentacijomis išplėstų duomenų variacijomis. Geriausias *top-k* metrikas pavyko pasiekti su modeliu, kurio duomenų rinkiniui buvo pritaikytos įvairesnės augmentacijos. Šiam modeliui pasiektas 76.10% tikslumas, 67.14% preciziškumas, 63.58% jautrumas ir 64.57% F1 (žr. **18 lentelė**). Šio tipo duomenų išplėtimas nepadėjo modeliui geriau apsimokyti, o priešingai – neigiamai paveikė modelio rezultatus. Šie modeliai nusileidžia visomis metrikomis: *top-k*, svorinėms (žr. **17 lentelė**) ir nesvorinėms (**18 lentelė**) tikslumo metrikoms, lyginant su modeliu, kuris buvo apmokytas su tuo pačiu, bet nepraplėstu duomenų rinkiniu. Tai reiškia, kad suprastėjo ne vien modelio gebėjimas generalizuoti klases, bet ir tiksliai jas atpažinti.

16 lentelė. Linijinio modelio su augmentuotais duomenimis rezultatai

Modelis	top-1 (%)	top-3 (%)	top-5 (%)
Linijinis modelis su augmentuotais duomenimis (1) (<i>RandomVerticalFlip(p=0.5)</i> , <i>RandomHorizontalFlip(p=0.5)</i> , <i>RandomRotation(degrees=30)</i> , <i>RandomResizedCrop(size=256, scale=(0.8, 1.0))</i> , <i>ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.2)</i> , <i>GaussianBlur(kernel_size=5, sigma=(0.1, 2.0))</i>)	76.10	92.72	96.94
Linijinis modelis su augmentuotais duomenimis (2) (<i>RandomVerticalFlip(p=0.5)</i> , <i>RandomHorizontalFlip(p=0.5)</i> , <i>ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.2)</i>)	75.68	92.48	96.64
Linijinis modelis su augmentuotais duomenimis (3) (<i>RandomVerticalFlip(p=0.5)</i> , <i>RandomHorizontalFlip(p=0.5)</i>)	75.71	92.54	96.78

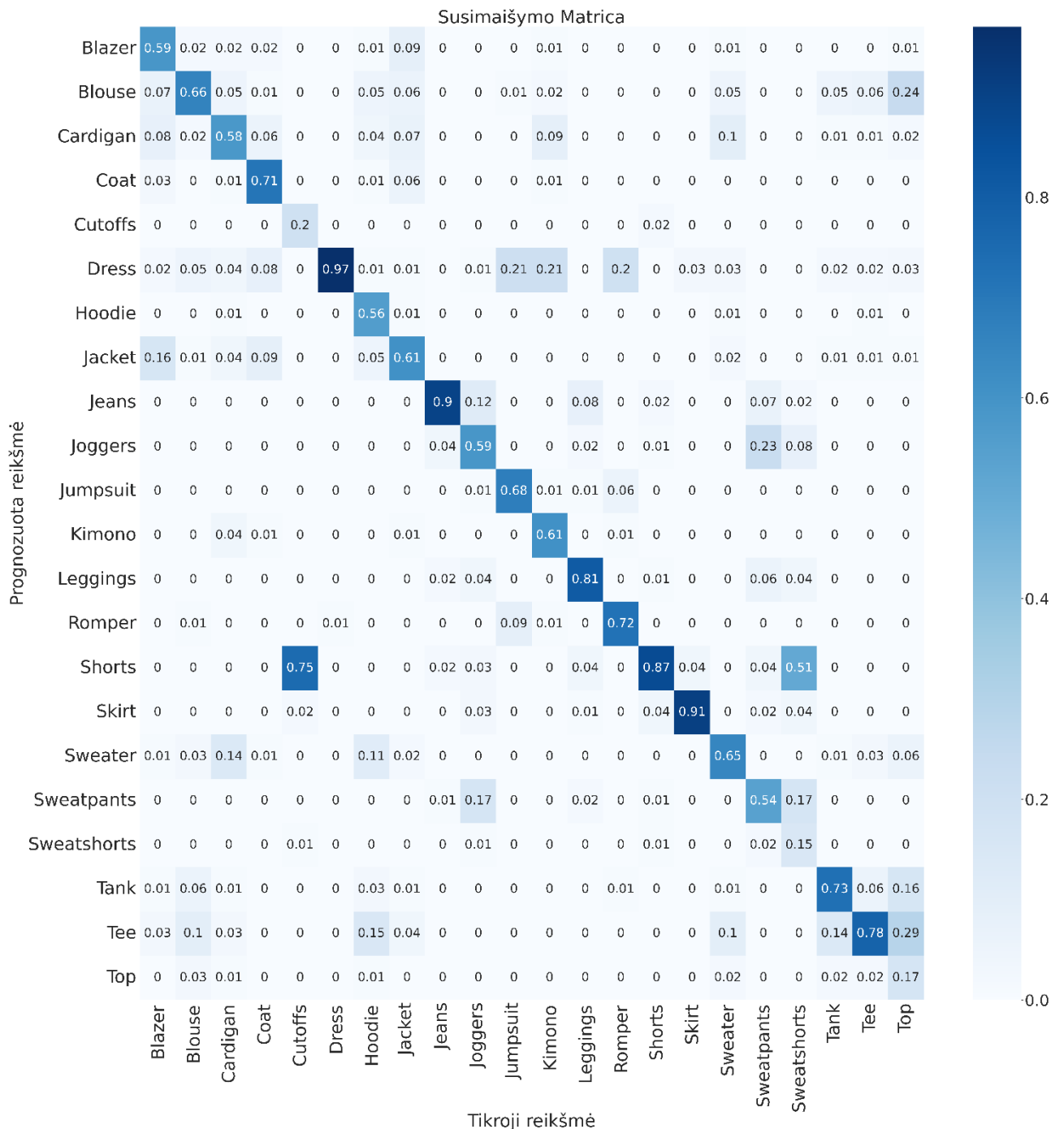
17 lentelė. Linijinio modelio su augmentuotais duomenimis svoriniai tikslumo rezultatai

Modelis	Tikslumas (Svorinis)	Preciziškumas (Svorinis)	Jautrumas (Svorinis)	F1 (Svorinis)
Linijinis modelis su augmentuotais duomenimis (1)	86.94	74.97	76.10	75.13
Linijinis modelis su augmentuotais duomenimis (2)	85.99	75.16	75.68	75.27
Linijinis modelis su augmentuotais duomenimis (3)	85.75	75.37	75.71	75.38

18 lentelė. Linijinio modelio su augmentuotais duomenimis makro tikslumo rezultatai

Modelis	Tikslumas (makro)	Preciziškumas (makro)	Jautrumas (makro)	F1 (makro)
Linijinis modelis su augmentuotais duomenimis (1)	76.10	67.14	63.58	64.57
Linijinis modelis su augmentuotais duomenimis (2)	75.68	65.99	65.74	65.57
Linijinis modelis su augmentuotais duomenimis (3)	75.71	65.74	66.43	65.68

Žemiau pateiktoje susimaišymo matricoje matomas susimaišymo padidėjimas, lyginant su modeliu, apmokytu su apjungtomis klasėmis (žr 30 pav.). Sumažėjo *Cutoffs*, *Leggings* klasių susimaišymas. Padidėjo *Blazer*, *Blouse*, *Cardigan*, *Coat*, *Hoodie*, *Jacket*, *Jeans*, *Joggers*, *Jumpsuit*, *Kimono*, *Romper*, *Shorts*, *Sweater*, *Sweatpants*, *Sweatshorts*, *Tank*, *Top* klasių susimaišymas. Nepasikeitė *Dress*, *Skirt*, *Tee* klasių susimaišymas. Tarp nepakitusių klasių – didelį kiekį paveikslėlių turinčios klasės.



30 pav. Linijinio modelio su augmentuotais duomenimis rezultatai

3.5. Bendri rezultatai

Toliau pateikiami bendri eksperimentų rezultatai ir jų palyginimas. Metrikos apskaičiuotos panaudojant testavimo duomenų rinkinį.

Žemiau pateiktoje lentelėje (žr. **19 lentelė**) pateikiamos eksperimentų modelių *top-1*, *top-3*, *top-5* metrikos. Literatūroje apžvelgtų modelių *top-1* metrikos nėra nurodytos, todėl jų palyginti nepavyks. *Top-k* metrikos lyginamos su esamais implementuotais *DeppFashion* duomenų rinkinio klasifikavimo užduoties sprendimo metodais. Tarp visų eksperimentų, geriausių rezultatų pasiekė ir pagerino esamas *top-3* ir *top-5* metrikas, taip pat geriausių tikslumo (*top-1*) metriką pasiekė linijinis modelis, apmokytas su *StyleGAN3* metodu išplėstu duomenų rinkiniu (77.53% *top-1*, 93.48% *top-3*, 97.40% *top-5*). Antroje vietoje pagal *top-3* ir *top-5* metrikas *StyleGAN2-ADA* išplėstu duomenų rinkiniu apmokytas modelis (77.30% *top-1*, 93.44% *top-3*, 97.34% *top-5*). Nors šio modelio *top-3* ir *top-5* metrikos yra antros pagal vertę, tačiau lyginant *top-1* vertes, antroje vietoje būtų linijinis modelis, apmokytas su nepraplėstu apjungtų klasių duomenų rinkiniu (77.33% *top-1*, 93.36% *top-3*, 97.26% *top-5*). Tačiau lyginant metrikas, atsižvelgiant į literatūroje pateikiamas, šis modelis yra tik trečias pagal rezultatus. Linijinis modelis su originaliomis klasėmis nors ir nepasiekė aukščiausių rezultatų, tačiau pagerino literatūroje apžvelgtus rezultatus (76.15% *top-1*, 92.59% *top-3*, 96.60% *top-5*). Hierarchiniai modeliai nepagerino esamų rezultatų. Trijų lygių hierarchinis modelis pasiekė 76.08% *top-1*, 88.78% *top-3*, 89.88% *top-5*. Trijų nepilnų lygių hierarchinis modelis pagerino buvusio modelio rezultatus su 76.53% *top-1*, 91.39% *top-3*, 94.73% *top-5*. Dviejų lygių hierarchinis modelis pagerino anksčiau minėtų hierarchinių modelių rezultatus su 76.53% *top-1*, 91.78% *top-3*, 95.40% *top-5*. Linijinis modelis, apmokytas išplėtus duomenų rinkinį su augmenteduotais duomenimis pasiekė geresnius rezultatus nei hierarchiniai modeliai, tačiau nepagerino GAN metodais išplėtais duomenų rinkiniais apmokyty modelių (76.10% *top-1*, 92.72% *top-3*, 96.94% *top-5*).

19 lentelė. Bendros modelių top-k metrikos

Modelis	<i>top-1</i> (%)	<i>top-3</i> (%)	<i>top-5</i> (%)
Liu et al.	-	82.58	90.17
Lee et al.	-	91.37	95.26
Wang et al.	-	90.99	95.78
Shajini et al.	-	91.02	96.2
Shajini et al.	-	91.06	96.35
Zhang et al.	-	91.99	96.44
Linijinis modelis su originaliomis klasėmis	76.15	92.59	96.60
Linijinis modelis su apjungtomis klasėmis	77.33	93.36	97.26
Trijų lygių hierarchinis modelis	76.08	88.78	89.88
Trijų nepilnų lygių hierarchinis modelis	76.53	91.39	94.73
Dviejų lygių hierarchinis modelis	76.53	91.78	95.40
Linijinis modelis su <i>StyleGAN2-ADA</i> generuotais duomenimis	77.30	93.44	97.34
Linijinis modelis su <i>StyleGAN3</i> generuotais duomenimis	77.53	93.48	97.40
Linijinis modelis su augmenteduotais duomenimis	76.10	92.72	96.94

Žemiau pateiktoje lentelėje (žr. **20 lentelė**) pateiktos modelių svorinės rezultatų metrikos. Tarp šių rezultatų linijinis modelis su *StyleGAN3* generuotais duomenimis taip pat pasiekė geriausius rezultatus (tikslumas 87.49%, preciziškumas 76.71%, jautrumas 77.53%, *F1* 76.69%). Linijinis modelis su *StyleGAN2-ADA* generuotais duomenimis (tikslumas 87.43%, preciziškumas 76.35%, jautrumas 77.30%, *F1* 76.45%) pasiekė gana artimus linijinio modelio su apjungtomis klasėmis rezultatus (tikslumas 87.44%, preciziškumas 76.48%, jautrumas 77.33%, *F1* 76.55%), tačiau jų nepagerino.

20 lentelė. Bendros modelių svorinės tikslumo metrikos

Modelis	Tikslumas (Svorinis)	Preciziškumas (Svorinis)	Jautrumas (Svorinis)	<i>F1</i> (Svorinis)
Linijinis modelis su originaliomis klasėmis	87.45	75.22	76.15	75.21
Linijinis modelis su apjungtomis klasėmis	87.44	76.48	77.33	76.55
Trijų lygių hierarchinis modelis	86.57	75.24	76.08	75.43
Trijų nepilnų lygių hierarchinis modelis	86.86	75.79	76.53	76.00
Dviejų lygių hierarchinis modelis	86.79	75.79	76.53	76.03
Linijinis modelis su <i>StyleGAN2-ADA</i> generuotais duomenimis	87.43	76.35	77.30	76.45
Linijinis modelis su <i>StyleGAN3</i> generuotais duomenimis	87.49	76.71	77.53	76.69
Linijinis modelis su augmentuotais duomenimis	86.94	74.97	76.10	75.13

Žemiau pateiktoje lentelėje (žr. **21 lentelė**) pateiktos modelių makro rezultatų metrikos. Pagal tikslumą ir preciziškumą, geriausi rezultatai pasiekti linijinio modelio su *StyleGAN3* generuotais duomenimis (tikslumas 77.53%, preciziškumas 70.14%, jautrumas 65.19%, *F1* 66.36%). Tačiau pagal jautrumo ir *F1* metrikas, geriau pasirodė linijinis modelis su apjungtomis klasėmis (tikslumas 77.33%, preciziškumas 68.98%, jautrumas 66.33%, *F1* 66.90%). Modelių jautrumo metrikomis taip pat aplenkė ir modelis, apmokytas su *StyleGAN2-ADA* generuotais duomenimis (tikslumas 77.30%, preciziškumas 68.95%, jautrumas 65.43%, *F1* 66.15%). Augmentacijomis išplėsto duomenų rinkinio modelis ir hierarchiniai modeliai pasiekė žemesnius rezultatus nei GAN metodais išplėstais duomenų rinkiniais apmokyti modeliai.

21 lentelė. Bendros modelių makro tikslumo metrikos

Modelis	Tikslumas (makro)	Preciziškumas (makro)	Jautrumas (makro)	<i>F1</i> (makro)
Linijinis modelis su originaliomis klasėmis	76.15	60.96	47.91	49.01
Linijinis modelis su apjungtomis klasėmis	77.33	68.98	66.33	66.90
Trijų lygių hierarchinis modelis	76.08	67.45	63.49	64.72
Trijų nepilnų lygių hierarchinis modelis	76.53	67.50	65.10	65.72
Dviejų lygių hierarchinis modelis	76.53	67.05	65.18	65.76
Linijinis modelis su <i>StyleGAN2-ADA</i> generuotais duomenimis	77.30	68.95	65.43	66.15

Linijinis modelis su <i>StyleGAN3</i> generuotais duomenimis	77.53	70.14	65.19	66.36
Linijinis modelis su augmentuotais duomenimis	76.10	67.14	63.58	64.57

Išvados

1. Literatūros analizės metu pastebėta, kad drabužių klasifikavimo užduotis yra plačiai sprendžiama, naudojant tiek žemos, tiek aukštos rezoliucijos duomenų rinkinius ir įvairius kompiuterinės regos metodus. Plačiai paplitęs *DeepFashion* duomenų rinkinys, kuris yra didžiausias viešai prieinamas duomenų rinkinys su mada susijusioms užduotims spręsti.
2. Atlikta literatūros analizė, kurios metu pastebėta, kad *DeepFashion* duomenų rinkinio klasifikacijos užduočiai spręsti dažnai pasitelkiamos modelių architektūros, kurios apmokomos naudojant ne vien drabužių nuotraukas, bet ir orientacinius drabužių taškų duomenis.
3. Sukurta dviejų lygių hierarchinio klasifikavimo metodika leido pasiekti 76.53% *top-1*, 91.78% *top-3*, 95.40% *top-5* tikslumą nesubalansuotame *DeepFashion* duomenų rinkinyje. Tai rodo, kad tiesinis modelis leidžia efektyviau apdoroti multi-klases klasifikavimo užduotis nei hierarchinė architektūra.
4. Eksperimentai parodė, kad klasių disbalansas daro reikšmingą įtaką klasifikavimo tikslumui, o mažųjų klasių praplėtimas aukštos kokybės generuotais vaizdais leidžia pasiekti geresnį modelio apibendrinimą nei naudojant vien standartines augmentacijos technikas.
5. Pasiūlyta duomenų rinkinio balansavimo metodika, pasitelkiant *StyleGAN3* generuotus sintetinius vaizdus, padėjo ženkliai sumažinti duomenų disbalanso poveikį. Su GAN praplėstu duomenų rinkiniu modelio tikslumas išaugo iki 77,53% *top-1*, 93,48% *top-3* ir 97,40% *top-5*, kas viršijo tiek literatūroje aprašytus rezultatus, tiek kitus naudotus balansavimo metodus.
6. *StyleGAN3* modelio taikymas buvo efektyvesnis nei *StyleGAN2-ADA*, kadangi šis metodas greičiau konvergavo ir generavo mažiau artefaktų, taip užtikrinant aukštesnę sugeneruotų duomenų kokybę ir geresnius klasifikavimo rezultatus.

Literatūros sąrašas

- [1] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, “DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 1096–1104. doi: 10.1109/CVPR.2016.124.
- [2] S. Sadegharmaki, M. A. Kastner, and S. Shin’ichi, “FashionGraph: Understanding fashion data using scene graph generation,” in *Proceedings - International Conference on Pattern Recognition*, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 7923–7929. doi: 10.1109/ICPR48806.2021.9412662.
- [3] C. Huang, Y. Li, C. C. Loy, and X. Tang, “Learning deep representation for imbalanced classification,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 5375–5384. doi: 10.1109/CVPR.2016.580.
- [4] W. Ouyang, X. Wang, C. Zhang, and X. Yang, “Factors in Finetuning Deep Model for object detection,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 864–873, 2016, doi: <https://doi.org/10.48550/arXiv.1601.05150>.
- [5] G. Van Horn and P. Perona, “The Devil is in the Tails: Fine-grained Classification in the Wild,” Sep. 2017, doi: 10.48550/arXiv.1709.01450.
- [6] J. Kozerawski, V. Fragoso, N. Karianakis, G. Mittal, M. Turk, and M. Chen, “BLT: Balancing Long-Tailed Datasets with Adversarially-Perturbed Images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 338–355. doi: 10.1007/978-3-030-69535-4_21.
- [7] R. Sauber-Cole and T. M. Khoshgoftaar, “The use of generative adversarial networks to alleviate class imbalance in tabular data: a survey,” *J Big Data*, vol. 9, Dec. 2022, doi: 10.1186/s40537-022-00648-6.
- [8] S. Cusworth, G. V. Gkoutos, and A. Acharjee, “A novel generative adversarial networks modelling for the class imbalance problem in high dimensional omics data,” *BMC Med Inform Decis Mak*, vol. 24, Dec. 2024, doi: 10.1186/s12911-024-02487-2.
- [9] N. Chereddy and B. K. Bolla, “Evaluating the Utility of GAN Generated Synthetic Tabular Data for Class Balancing and Low Resource Settings,” Jun. 2023, doi: 10.1007/978-3-031-36402-0_4.
- [10] P. Yadav, G. Sihag, and V. Vijay, “Rebalancing the Scales: A Systematic Mapping Study of Generative Adversarial Networks (GANs) in Addressing Data Imbalance,” Feb. 2025, doi: 10.48550/arXiv.2502.16535.
- [11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR, 2015.
- [12] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 483–499. doi: 10.1007/978-3-319-46484-8_29.

- [13] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans Pattern Anal Mach Intell*, vol. 39, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.
- [14] W. Wang, Y. Xu, J. Shen, and S. C. Zhu, “Attentive Fashion Grammar Network for Fashion Landmark Detection and Clothing Category Classification,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2018, pp. 4271–4280. doi: 10.1109/CVPR.2018.00449.
- [15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Int J Comput Vis*, vol. 128, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [16] S. Lee, H. Eun, S. Oh, W. Kim, C. Jung, and C. Kim, “Landmark-free clothes recognition with a two-branch feature selective network,” *Electron Lett*, vol. 55, pp. 745–747, Jun. 2019, doi: 10.1049/el.2019.0660.
- [17] P. Ballester and R. M. Araujo, “On the performance of googlenet and alexnet applied to sketches,” in *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, AAAI press, 2016, pp. 1124–1128. doi: 10.1609/aaai.v30i1.10171.
- [18] L. A. Gatys, A. S. Ecker, and M. Bethge, “Texture synthesis using convolutional neural networks,” in *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, 2015, pp. 262–270.
- [19] R. Geirhos, C. Michaelis, F. A. Wichmann, P. Rubisch, M. Bethge, and W. Brendel, “Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness,” in *7th International Conference on Learning Representations, ICLR 2019*, International Conference on Learning Representations, ICLR, 2019.
- [20] Y. Zhang, P. Zhang, C. Yuan, and Z. Wang, “Texture and shape biased two-stream networks for clothing classification and attribute recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2020, pp. 13535–13544. doi: 10.1109/CVPR42600.2020.01355.
- [21] M. Shajini and A. Ramanan, “An improved landmark-driven and spatial–channel attentive convolutional neural network for fashion clothes classification,” *Visual Computer*, vol. 37, pp. 1517–1526, Jun. 2021, doi: 10.1007/s00371-020-01885-7.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [23] S. Guo *et al.*, “The imaterialist fashion attribute dataset,” in *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 3113–3116. doi: 10.1109/ICCVW.2019.00377.
- [24] B. Loni, L. Y. Cheung, M. Riegler, A. Bozzon, L. Gottlieb, and M. Larson, “Fashion 10000: An enriched social image dataset for fashion and clothing,” in *Proceedings of the 5th ACM Multimedia Systems Conference, MMSys 2014*, Association for Computing Machinery, 2014, pp. 41–46. doi: 10.1145/2557642.2563675.
- [25] M. Shajini and A. Ramanan, “A knowledge-sharing semi-supervised approach for fashion clothes classification and attribute prediction,” *Visual Computer*, vol. 38, pp. 3551–3561, Nov. 2022, doi: 10.1007/s00371-021-02178-3.
- [26] S. K. Noh, “Recycled clothing classification system using intelligent IoT and deep learning with ALExNet,” *Comput Intell Neurosci*, vol. 2021, 2021, doi: 10.1155/2021/5544784.

- [27] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. van Gool, “Apparel classification with style,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7727 LNCS, no. PART 4, pp. 321–335, 2013, doi: 10.1007/978-3-642-37447-0_25/COVER.
- [28] MK Gurucharan, “Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network | upGrad blog.” Accessed: Jan. 16, 2023. [Online]. Available: <https://www.upgrad.com/blog/basic-cnn-architecture/>
- [29] D. Bhatt *et al.*, “CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope,” *Electronics 2021, Vol. 10, Page 2470*, vol. 10, no. 20, p. 2470, Oct. 2021, doi: 10.3390/ELECTRONICS10202470.
- [30] Sai Balaji, “Binary Image classifier CNN using TensorFlow | by Sai Balaji | Techiepedia | Medium,” Techiepedia. Accessed: Jan. 16, 2023. [Online]. Available: <https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697>
- [31] H. Li, R. Zhao, and X. Wang, “Highly Efficient Forward and Backward Propagation of Convolutional Neural Networks for Pixelwise Classification,” Dec. 2014, doi: <https://doi.org/10.3390/electronics12051199>.
- [32] Mrinal Tyagi, “Image Segmentation: Part 1. Mathematical and practical... | by Mrinal Tyagi | Towards Data Science,” Towards Data Science. Accessed: Jan. 16, 2023. [Online]. Available: <https://towardsdatascience.com/image-segmentation-part-1-9f3db1ac1c50>
- [33] Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah, and Mohammed Bennamoun, “A Guide to Convolutional Neural Networks for Computer Vision | Springer eBooks | IEEE Xplore.” Accessed: Jan. 16, 2023. [Online]. Available: <https://ieeexplore.ieee.org/book/8295029>
- [34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125). ,doi.org/10.48550/arXiv.1612.03144,” 2016, doi: <https://doi.org/10.48550/arXiv.1612.03144>.
- [36] J. Terven and D. Cordova-Esparza, “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS,” Apr. 2023, doi: 10.3390/make5040083.
- [37] Ultralytics, “YOLO11 🚀 NEW,” <https://docs.ultralytics.com/models/yolo11/>.
- [38] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” Oct. 2024, doi: 10.48550/arXiv.2410.17725.
- [39] Wawan Kurniawan, A. Kurniasih, and Muhamad Abdul Ghani, “Real or Deepfake Face Detection in Images and Video Data using YOLO11 Algorithm,” *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, pp. 1514–1521, Feb. 2025, doi: 10.59934/jaiea.v4i2.939.
- [40] Ian J. Goodfellow *et al.*, “Generative Adversarial Networks,” *Adv Neural Inf Process Syst*, pp. 2672–2680, 2014.

- [41] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” *IEEE Trans Pattern Anal Mach Intell*, vol. 43, pp. 4217–4228, Dec. 2021, doi: 10.1109/TPAMI.2020.2970919.
- [42] X. Huang and S. Belongie, “Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization,” in *Proceedings of the IEEE International Conference on Computer Vision*, Institute of Electrical and Electronics Engineers Inc., Dec. 2017, pp. 1510–1519. doi: 10.1109/ICCV.2017.167.
- [43] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2020, pp. 8107–8116. doi: 10.1109/CVPR42600.2020.00813.
- [44] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, “Training generative adversarial networks with limited data,” in *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, 2020.
- [45] J. Joo *et al.*, “Enhancing automated strabismus classification with limited data: Data augmentation using StyleGAN2-ADA,” *PLoS One*, vol. 19, May 2024, doi: 10.1371/journal.pone.0303355.
- [46] A. Mohanty, A. Sutherland, M. Bezbradica, and H. Javidnia, “High-Fidelity Synthetic Face Generation for Rosacea Skin Condition from Limited Data,” *Electronics (Switzerland)*, vol. 13, Jan. 2024, doi: 10.3390/electronics13020395.
- [47] M. K. Woodland *et al.*, “Evaluating the Performance of StyleGAN2-ADA on Medical Images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 142–153. doi: 10.1007/978-3-031-16980-9_14.
- [48] R. Chen *et al.*, “Decoupled Data Augmentation for Improving Image Classification,” Oct. 2024, doi: 10.48550/arXiv.2411.02592.

Priedai

1 Priedas. Pristatytas mokslinis straipsnis 30-ojoje tarptautinėje informacinių technologijų konferencijoje „IVUS 2025“ Moleikaitytė A., Paulauskas J. (2025). Improving DeepFashion Dataset Classification Accuracy with StyleGAN2-ADA: Addressing Imbalanced Data in Fashion Image Recognition.

2 Priedas. Dalyvavimo „IVUS 2025“ konferencijoje sertifikatas.

3 Priedas. M. Bacevicius, A. Paulauskaite-Taraseviciene, G. Zokaityte, L. Kersys, and A. Moleikaityte, “Comparative Analysis of Perturbation Techniques in LIME for Intrusion Detection Enhancement,” Mach Learn Knowl Extr, vol. 7, Mar. 2025, doi: 10.3390/make7010021.