

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
INFORMATIKOS KATEDRA

**Kompiuterinių žaidimų amžiaus cenzo nustatymas  
naudojantis konvoliuciniais neuroniniais tinklais**

**Age Rating for Video Games Using Convolutional Neural Networks**

Magistro baigiamasis darbas

Atliko: Laurynas Lazauskas (parašas)

Darbo vadovas: dr. Olga Kurasova (parašas)

Recenzentas: lekt. Irmantas Radavičius (parašas)

Vilnius – 2021

## Santrauka

Nuolatinis kompiuterinių žaidimų industrijos augimas kelia poreikį automatizuotam žaidimų turinio kuravimui ir amžiaus cenzo kategorijos nustatymui. Literatūroje tokio uždavinio sprendimų nerasta. Šiame darbe minėtas uždavinys sprendžiamas pritaikant giliuosius neuroninius tinklus.

Uždaviniui spręsti sukurta kompiuterinių žaidimų duomenų aibė, kurią sudaro tūkstantis žaidimų, 11 tūkst. vaizdo įrašų, tekstiniai aprašai, PEGI ir ESRB įvertinimai ir kiti meta duomenys.

Išnagrinėjus literatūrą, kur sprendžiami panašūs uždaviniai (smurto paieška vaizduose, ginklų šūvio garso atpažinimas ir pan.), sukurta daugiau nei 12 tinklų spręsti šį uždavinį. Geriausi tinklai buvo komponuojami skirtingais būdais, taip tinklai amžiaus cenzo kategoriją nustatė atsižvelgdami į visus prieinamus duomenis – vaizdo įrašo kadrus, judesį, garsą ir tekstinius žaidimo aprašus.

Komponuoti tinklai turėjo geresnį tikslumą nei pavieniai: vieni pasižymėjo geru vaikų (93 %) ir suaugusių (81 %) žaidimų klasifikavimo tikslumu, kiti pasiekė geriausią bendrą tikslumą (63 %).

Uždavinys pasirodė sudėtingas. Ypač sunku atpažinti paaugliams skirtus žaidimus, dėl jų panašumo į kitas kategorijas. Taip pat sudėtinga dėl duomenų gausos ir mažo požymių tankio: vaizdo įrašai ilgi, tačiau reikalinga informacija matoma ne visada. Šiuo atveju tekstiniai duomenys turi labai didelį požymių tankį ir parodo didžiausią tikslumą lyginant su vaizdo ar garso duomenimis atskirai.

Raktiniai žodžiai: kompiuterinių žaidimų amžiaus cenzo nustatymas, dirbtiniai neuroniniai tinklai, daugiarūšiai neuroniniai tinklai.

## Summary

The increasing growth in video games industry raises the need for automated content curation and age rating. No solutions for this problem were found in the literature. In this work we attempt to solve the problem using deep neural networks.

A data set for this problem was created. It contains over a thousand different video games, 11 thousand video excerpts, text descriptions, PEGI and ESRB ratings and other meta data.

Based on the literature solving similar classification problems (violence detection in videos, gunshot sound detection in sound, etc.), we created more than 12 neural networks. Networks showing the best accuracy were further composed using different methods, thus making classification consider all available data – video frames, optical flows, sounds and text descriptions.

Composed networks showed better accuracy compared to individual ones: one was good at classifying children (93 %) and adult (81 %) games, the other had the highest overall accuracy (63 %).

The problem was found to be difficult. The most difficult part was to classify games rated for teenagers, that is due to how similar they are either to children or adult games. The other difficult part was the abundance of data and low saturation of relevant information: videos are long, but only a part of them contains important information. Text descriptions proved to be superior compared to other data types in this problem, due to their high density of important information they showed the best accuracy among individual data type networks.

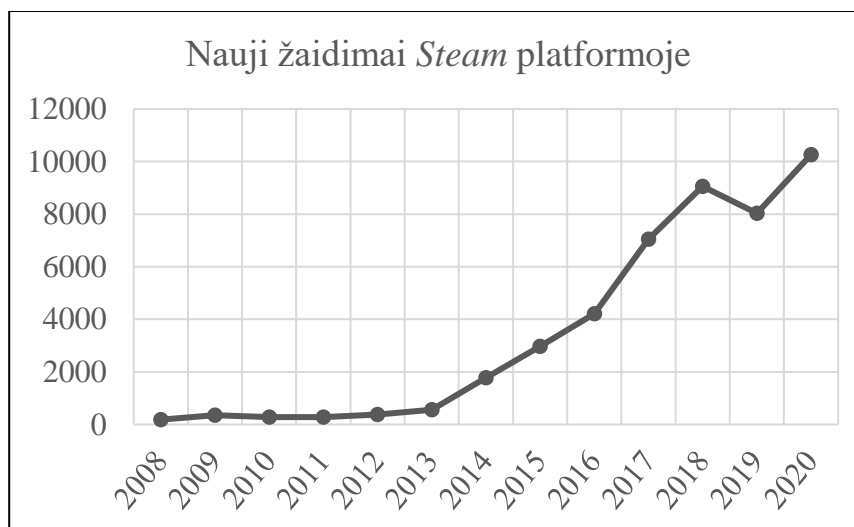
Keywords: video games age rating, artificial neural networks, multimodal neural networks.

# Turinys

ĮVADAS.....	1
1. ŽAIDIMŲ AMŽIAUS CENZO VERTINIMO SISTEMOS .....	5
1.1. PEGI .....	5
1.2. ESRB .....	6
1.2.1. ESRB kategorijos nustatymas naudojantis žaidimo požymiais.....	6
1.3. Apibendrinimas.....	7
2. AMŽIAUS KATEGORIJŲ POŽYMIUS KLASIFIKUOJANČIŲ TINKLŲ ANALIZĖ .....	8
2.1. Vaizdiniai duomenys .....	8
2.1.1. Konvoliuciniai neuroniniai tinklai.....	8
2.1.2. Video medžiagos įvestis .....	10
2.2. Garsiniai duomenys .....	15
2.3. Tekstiniai duomenys .....	17
2.4. Daugiarūšės tinklų architektūros.....	18
3. DUOMENŲ AIBĖS SUKŪRIMAS .....	20
3.1. Žaidimų sąrašo sudarymas.....	20
3.2. Vaizdinė medžiaga.....	22
3.3. Garso medžiaga.....	22
3.4. Tekstiniai aprašai .....	23
3.5. Duomenų kokybė.....	23
3.6. Apibendrinimas.....	24
4. NEURONINIŲ TINKLŲ KŪRIMAS IR MOKYMAS.....	25
4.1. Darbas su statine vaizdine medžiaga .....	25
4.1.1. Paprasti konvoliuciniai sluoksniai.....	25
4.1.2. Iš anksto apmokyti konvoliuciniai sluoksniai .....	27
4.2. Darbas su dinamine vaizdine medžiaga.....	28
4.2.1. Trijų dimensijų konvoliuciniai filtrai .....	29
4.2.2. Konvoliuciniai filtrai su ilga trumpalaike atmintimi .....	30
4.2.3. Išretinti kadrai ir optiniai judesio srautai.....	31
4.3. Darbas su garsine medžiaga.....	34
4.3.1. Spektrogramos ir konvoliuciniai filtrai .....	34
4.3.2. Iš anksto apmokyti neuroniniai tinklai .....	35
4.4. Darbas su tekstine medžiaga.....	35
4.4.1. Žodžių rinkinys.....	36
4.4.2. Žodžių įterpiniai .....	38
4.5. Apibendrinimas.....	40
5. NEURONINIŲ TINKLŲ KOMPONAVIMAS.....	41
5.1. Kelių tinklų rezultatų apjungimas funkcija.....	41
5.2. Daugiarūšė komponuota tinklo architektūra.....	46
5.3. Pakartotinė multimedijos įvestis.....	49
REZULTATAI IR IŠVADOS.....	52
LITERATŪRA.....	54

## Įvadas

Kompiuterinių žaidimų industrija nuolatos auga. Taip pat matomas nuolatinis sukuriamų ir internetinėje erdvėje platinamų žaidimų skaičiaus augimas. Pagal *Statista* pateikiamus duomenis<sup>1</sup>, vienoje populiariausių kompiuterinių žaidimų pardavimų platformoje *Steam*, per 2020 metus, kiekvieną dieną vidutiniškai buvo išleidžiami 28 žaidimai. Kaip matome 1 pav. tai yra žymiai daugiau naujų žaidimų nei prieš kelis metus. Kitoje platformoje *Itch.io*, dar 2017 metais<sup>2</sup>, kiekvieną dieną vidutiniškai buvo išleidžiami 187 žaidimai.



1 pav. Naujai išleistų žaidimų skaičius Steam platformoje kiekvienais metais tarp 2008 ir 2020

Didelis publikuojamų žaidimų skaičius sukuria turinio kuravimo problema, t. y. atsakingiems asmenims darosi neįmanoma peržiūrėti visų žaidimų turinį ir įvertinti, kokioms amžiaus grupėms jis yra tinkamas, todėl atsiranda automatizuoto sprendimo poreikis.

Automatizuotas žaidimo amžiaus cenzo nustatymas yra sudėtingas klasifikavimo uždavinys. Sprendimas turėtų išanalizuoti apie žaidimą prieinamus duomenis ir jais remdamasis priskirti vieną iš galimų amžiaus kategorijų. Dažniausiai apie žaidimus prieinami duomenys yra tekstiniai aprašai, žaidimo eigoje (angl. *gameplay*) užfiksuoti statiniai ir dinaminiai vaizdai su garsu.

Klasifikavimo tikslumas stipriai priklauso nuo pateikiamų duomenų reprezentatyvumo. Standartiškai amžiaus cenzo vertinamas atliekamas atsižvelgiant į visą žaidimo turinį, t. y. vertinamas

---

<sup>1</sup> *Statista* ataskaitos apie *Steam* platformoje išleidžiamus žaidimus  
<https://www.statista.com/statistics/552623/number-games-released-steam/>.

<sup>2</sup> *Itch.io* apžvalginė 2017 metų ataskaita <https://itch.io/blog/19998/itchio-year-in-review-2017>.

visas žaidimas nuo pradžios iki pabaigos. Automatinio klasifikavimo atveju nagrinėti visą turinį gali būti neįmanoma dėl resursų ir tokių duomenų trūkumo. Tačiau, jeigu nagrinėjamuose duomenyse pateikiamos ištraukos atitinka visą žaidimą, kaip, pvz., žaidimo anonsuose, tai klasifikavimas turėtų būti įmanomas.

Amžiaus cenzo kategorijos skirstomos ir vertinamos įvairiai. Pavyzdžiui, Europos vaizdo žaidimų turinio vertinimo sistema (angl. *Pan European Game Information*, toliau PEGI) amžių skirsto į 5 grupes: 3+, 7+, 12+, 16+ ir 18+ metų. Amžiaus grupė žaidimui nustatoma pagal tam tikrus požymius, pvz., žaidimas priskiriamas į 18+ amžiaus grupę, jeigu jame yra vaizduojamas šiurkštus smurtas, seksualinė veikla arba vulgari kalba. Taigi, automatizuotas sprendimas turėtų žinoti kiekvienai amžiaus grupei būdingus požymius ir mokėti juos tiksliai identifikuoti turimuose duomenyse. Dėl požymių ir požymių išraiškų gausos tokių požymių rinkinį sudaryti yra labai sudėtinga.

Dirbtiniai neuroniniai tinklai (angl. *artificial neural networks*, toliau DNT) ir ypač gilieji neuroniniai tinklai (angl. *deep neural networks*, toliau GNT) rodo puikius rezultatus sprendžiant sudėtingus klasifikavimo uždavinius. Klasifikavimo uždavinius sprendžiantys neuroniniai tinklai kaip įvestį priima norimus klasifikuoti duomenis skaitiniu formatu, o kaip išvestį pateikia rinkinį tikimybių, kurios nurodo ar pateikti duomenys priklauso atitinkamoms klasėms. Svarbus GNT privalumas yra didelis kiekis skaičiavimo sluoksnių, kurie perduoda abstrakčius skaičiavimo rezultatus iš vieno sluoksnio į kitą ir taip leidžia tinklui automatiškai išmokti išskirti reikalingus duomenų požymius ir juos identifikuoti pateiktuose duomenyse. Tai stipriai palengvina klasifikavimo uždavinį, kadangi nereikia rankiniu būdu sudarinėti požymių rinkinių ar ieškoti sprendimų kaip požymius atpažinti turimuose duomenyse. Duomenų požymius išrinkti ir duomenis klasifikuoti tinklas išmoksta mokymosi su mokytoju (angl. *supervised learning*) metu, kai jam yra pateikiamas didelis kiekis iš anksto teisingai suklasifikuotų duomenų [Sch14, LBH15].

Nors GNT puikiai sprendžia klasifikavimo uždavinius ir nereikalauja daug rankinio darbo, jų taikymas automatiniam kompiuterinių žaidimų amžiaus cenzo nustatymui nėra trivialus. Iškyla daug klausimų, pvz., kokius įvesties duomenis naudoti, kokiame formate juos pateikti, kokią tinklo architektūrą naudoti, ar tikslumas gali būti pakankamai geras.

Nagrinėjant literatūrą galima rasti panašius uždavinius sprendžiančių tinklų. Taip pat pastebima, kad tokiems uždaviniams spręsti naudojama įvairaus tipo duomenys. Taikant įvairias konvoliucinių tinklų architektūras yra apdorojama vaizdo medžiaga ir sėkmingai atpažįstamas smurtas ([SL17]) ir kitokie veiksmai ([NHV+15]). Iš garso spektrogramų sėkmingai atpažįstami ginklo šūviai ([CPH+17]), kalba ir jos akcentai. Iš tekstinių duomenų naudojantis efektyviais teksto įterpimo mechanizmais pagerinami žaidimų reitingą atspėjantys tinklai ([BBR+18]).

Šio darbo tikslas yra sukurti dirbtinį neuroninį tinklą siekiant kuo didesnio kompiuterinių žaidimų klasifikavimo pagal amžiaus cenzą tikslumo. Remiantis srities specialistais, kurie sprendžia panašias klasifikavimo užduotis, išrenkamos neuroninių tinklų architektūros, jos įgyvendinamos, palyginamas jų tarpusavio tikslumas. Analizuojamos ir įgyvendinamos skirtingų tinklų galimos kompozicijos ir modifikacijos siekiant kuo didesnio klasifikavimo tikslumo.

Šio darbo uždaviniai:

1. Identifikuoti kokia informacija apie žaidimą reikalinga siekiant nustatyti amžiaus cenzo kategoriją.
2. Surasti geriausius metodus, sprendžiančius panašius uždavinius pagal turimus duomenis ir siekiamus nustatyti amžiaus kategorijų požymius.
3. Surinkti ir paruošti kompiuterinių žaidimų multimedijos duomenų aibę, tinkamą DNT apmokymui sprendžiant amžiaus cenzo klasifikavimo problemą.
4. Kiekvienam duomenų tipui išrinkti tinklų architektūras su geriausiu tikslumu klasifikuojančias amžiaus cenzo kategorijas.
5. Komponuojant ir/ar modifikuojant esamas architektūras sukurti naujas DNT architektūras, pritaikytas gerai klasifikuoti žaidimus į amžiaus cenzo grupes.
6. Įgyvendinus tinklų architektūras, palyginti ir išrinkti kokios duomenų rūšys ar jų kombinacijos ir kokios architektūros leidžia gauti geriausius rezultatus.

Darbe yra penki pagrindiniai skyriai, po kurių seka rezultatai ir išvados.

Pirmame analizės skyriuje išnagrinėta probleminė sritis ir identifikuota kaip ir kuo remiantis amžiaus cenzo kategorijos nustatomos dvejose populiariose amžiaus cenzą vertinančiose sistemose. Abi sistemos palygintos tarpusavyje ir nustatyta, kad amžiaus kategorijų vertinimą jos supranta panašiai.

Antrame analizės skyriuje išnagrinėta literatūra, kurioje sprendžiami panašūs uždaviniai ir išrinkti tinkami komponentai spręsti kompiuterinių žaidimų amžiaus kategorijos klasifikavimo

uždavinį. Literatūros paieška remiasi pirmame skyriuje išrinktais amžiaus kategorijos nustatymui svarbiausiais požymiais. Taip pat atsižvelgiama į apie žaidimus prieinamų duomenų tipus. Išrinkti komponentai apima visus tris žaidimus apibūdinančius duomenų tipus: vaizdus, garsus, tekstinius aprašus ir jų kombinacijas.

Trečiame skyriuje aprašytas duomenų aibės sukūrimas. Pateiktas duomenims surinkti taikytas metodas, duomenų aibės turinys: duomenų tipai, kiekiai, formatai ir t. t.

Ketvirtame skyriuje aprašytas duomenų tipams specifinių neuroninių tinklų kūrimas ir analizavimas. Čia įgyvendintos ir įvertintos skirtingos strategijos dirbti su neuroniniais tinklais apdorojant statinius ir dinامينius vaizdus, garsus ir tekstinius aprašus.

Penktame skyriuje aprašyti skirtingi neuroninių tinklų komponavimai. Pateikti du būdai, kaip komponuoti tinklus: apjungiant skirtingų tinklų rezultatus apjungimo funkcija ir sujungiant kelis tinklus į vieną daugiarūšį tinklą naudojantis apjungimo sluoksniais.



## 1. Žaidimų amžiaus cenzo vertinimo sistemos

Žaidimai skirstomi į amžiaus kategorijas įvairiai. Kai kurios šalys ir regionai naudoja savo sukurtas vertinimo sistemas, kitos naudoja jau egzistuojančias vertinimo sistemas. Skirtingų sistemų vertinimas ir amžiaus kategorijos skiriasi, todėl, dėl vertinimo sistemų gausos, analizuoti visas būtų sudėtinga. Toliau bus apžvelgiamos dvi populiarios vertinimo sistemos: Europos vaizdo žaidimų turinio vertinimo sistema (angl. *Pan European Game Information*, toliau PEGI) ir JAV naudojama Pramoginių programų vertinimo taryba (angl. *Entertainment Software Rating Board*, toliau ESRB).

### 1.1. PEGI

PEGI vertinimo sistema išskiria du rezultatus: amžiaus kategoriją (angl. *age label*) ir turinio požymius (angl. *content descriptors*). Galimos amžiaus kategorijos yra: 3+, 7+, 12+, 16+ ir 18+ metų.

PEGI vertinimas yra iš dalies automatizuotas. Pirminis PEGI amžiaus kategorijos ir turinio požymių nustatymas vyksta automatiškai pagal standartinio klausimyno atsakymus. Tačiau galutinį sprendimą ir koregavimus atlieka ekspertų grupė, kruopščiai peržiūrėjusi vertinamo žaidimo turinį.

Nors galutinis vertinimo sprendimas paliekamas eksperto nuomonei, egzistuoja tam tikros gairės tarp žaidimo turinio požymių ir amžiaus kategorijų. Šios gairės buvo surinktos, išanalizavus PEGI vertinimo aprašus, ir pateiktos 1 lentelėje.

1 lentelė. Ryšiai tarp požymių ir amžiaus kategorijų. Pagal aprašus iš <https://pegi.info/what-do-the-labels-mean>.

Kategorijos Požymiai	3+	7+	12+	16+	18+
Smurtas (angl. <i>violence</i> )	Nėra	Nerealistiškas arba nerodantis detalių	Fantastinėje aplinkoje ir į žmones orientuotas tik nerealistiškas	Labiau realistiškas smurtas	Labiau realistiškas, šlykštus ar beprasmiškas smurtas
Vulgari kalba (angl. <i>bad language</i> )	Nėra	Nėra	Ne stipriai vulgari kalba	Seksualiniai keiksmažodžiai ir piktžodžiavimas	Seksualiniai keiksmažodžiai ir piktžodžiavimas
Baimė (angl. <i>fear</i> )	Nėra	Vaiką galintys išgąsdinti garsai ir vaizdai	Baisūs garsai ir siaubo efektai (be smurto)	Nėra	Nėra
Lošimai (angl. <i>gambling</i> )	Nėra	Nėra	Mokymas arba skatinimas lošti	Mokymas arba skatinimas lošti	Mokymas arba skatinimas lošti
Seksualinė veikla (angl. <i>sex</i> )	Nėra	Nėra	Seksualinės pozicijos ar užuominos	Erotinis nuogumas arba dalinai matoma seksualinė veikla	Aiškliai vaizduojama seksualinė veikla
Narkotikai (angl. <i>drugs</i> )	Nėra	Nėra	Nėra	Vaizduojamas narkotikų alkoholio ar tabako vartojimas	Išaukštinamas ir vaizduojamas narkotikų alkoholio ar tabako vartojimas

Diskriminacija (angl. <i>discrimination</i> )	Nėra	Nėra	Nėra	Nėra	Etniniai, religiniai, nacionalistiniai ar kitokie stereotipai skatinantys neapykanta
--	------	------	------	------	---

## 1.2. ESRB

ESRB vertinimo sistema išskiria tris vertinimo rezultatus: amžiaus kategoriją, žaidimo požymius (angl. *descriptors*) ir interaktyvius elementus (angl. *interactive elements*). Šiuo metu naudojamos 5-ios amžiaus kategorijos: visiems (angl. *everyone*), paaugliams (angl. *teen*) 13+, subrendusiems (angl. *mature*) 17+ ir suaugusiems (angl. *adults only*) 18+. Šiuo metu naudojama 30 skirtingų požymių, kurie grupuojami į 8 grupės: medžiagų vartojimas (angl. *substances*), kraujas (angl. *blood/gore*), smurtas (angl. *violence*), humoras (angl. *humor*), kalba (angl. *language*), nuogumas (angl. *nudity*), lošimai (angl. *gambling*), seksualumas (angl. *sexuality*). Interaktyvūs elementai apibūdina žaidėjo sąveiką su išoriniu pasauliu ir išskiriami tokie: pirkimai (angl. *in-game purchases*), atsitiktiniai pirkimai (angl. *in-game purchases including random items*), vartotojų bendravimas (angl. *users interact*), vietos dalinimasis (angl. *shares location*), neapribotas internetas (angl. *unrestricted internet*), internetinė, nevertinta muzika (angl. *online music not rated*).

Nors ESRB išskiria svarbius požymius, kuriais galima yra remiamasi nustatant amžiaus kategoriją, apibrėžtų ryšių tarp žaidimo požymių ir nustatomos amžiaus kategorijos nėra. Kaip ir PEGI, ESRB žaidimo turinį vertina ir galutinį vertinimo sprendimą priima nepriklausoma ekspertų grupė.

### 1.2.1. ESRB kategorijos nustatymas naudojantis žaidimo požymiais

[AAM+19] atliko tyrimą, kuriame išbandė ir palygino 6 skirtingus modelius, nustatančius ESRB amžiaus kategoriją naudojantis turimais žaidimo požymiais. Autoriai naudojo ESRB duomenų rinkinį, kurį sudarė 2053 skirtingi žaidimai, jų požymiai ir amžiaus kategorijos. Buvo palyginti tokie modeliai: bendrieji tiesiniai metodai (angl. *generalized linear models*), sprendimų medis (angl. *decision tree*), gilusis mokymasis (angl. *deep learning*), gradientiniai pagreitinami medžiai (angl. *gradient boosted trees*), atsitiktinis miškas (angl. *random forest*), naivus Bajeso metodas (angl. *Naive Bayes model*).

Tyrimo rezultatai parodė, kad galimas modelis, kuris naudodamas žaidimo požymius, nustato amžiaus kategoriją dideliu (90 %) tikslumu atitinkančią ekspertinę nuomonę. Visi modeliai pateikė panašiai gerus rezultatus, tačiau geriausiai tikslumo rezultatus parodė bendrieji tiesiniai metodai.

Taip pat buvo nustatyti didžiausią įtaką amžiaus kategorijai turintys požymiai. 2 lentelėje pateikti visi požymiai, išrikiuoti nuo labiausiai turinčio įtaką iki mažiausiai. Svoriai santykinai paskaičiuoti nuo labiausiai turinčio įtaką požymio. Autoriai išskyrė 8 požymius, kurie labiausiai įtakoja amžiaus kategorijos pasirinkimą (svoris >0,5): smurtas, stipriai vulgari kalba, fantastinis smurtas, kraujas, seksualinės temos, kraujas ir žiaurumas, stiprus smurtas, silpnas fantastinis smurtas.

2 lentelė. Požymiai įtakojantys amžiaus kategoriją ir jų santykiniai svoriai lyginant su labiausiai įtakingu

Požymis	Svoris	Požymis	Svoris
Smurtas	1	Animacinis smurtas	0,119
Stipriai vulgari kalba	0,838	Alkoholio ir tabako vartojimas	0,168
Fantastinis smurtas	0,795	Švelnus animacinis smurtas	0,159
Kraujas	0,756	Stipriai seksualinis turinys	0,154
Seksualinės temos	0,696	Narkotikų ir alkoholio vartojimas	0,154
Kraujas ir žiaurumas	0,681	Silpnas smurtas	0,134
Stiprus smurtas	0,663	Komiškos išdaigos	0,133
Silpnas fantastinis smurtas	0,599	Alkoholio rodymas	0,124
Vulgari kalba	0,389	Tabako vartojimas	0,107
Seksualiai provokuojančios temos	0,384	Animacinis kraujas	0,082
Silpnai vulgari kalba	0,338	Suaugusių humoras	0,063
Narkotikų rodymas	0,284	Vulgarūs dainų žodžiai	0,049
Šiurkštus humoras	0,241	Imitaciniai lošimai	0,017
Seksualinis turinys	0,236	Tabako rodymas	0,014
Narkotikų vartojimas	0,231	Seksualinis smurtas	0

### 1.3. Apibendrinimas

Nors skirtingos amžiaus cenzo vertinimo sistemos turi skirtingas amžiaus kategorijas, turinio požymiai yra panašūs. Palyginus PEGI ir ESRB išskiriamus turinio požymius matome, kad abi vertinimo sistemos turinyje ieško smurto, vulgarios kalbos, seksualinių elementų, priklausomybę keliančių medžiagų vartojimo ir lošimo elementų. PEGI taip pat išskiria diskriminaciją ir baime keliančius požymius, kurių ESRB nenaudoja.

Darbo [AAM+19] autoriai parodė, kad egzistuoja aiškus ryšys tarp žaidimo požymių ir žaidimo amžiaus kategorijos. Taip pat parodė, kad kai kurie požymiai yra žymiai svarbesni, nustatant amžiaus kategoriją, nei kiti. Ši informacija bus naudinga toliau sprendžiant amžiaus kategorijos klasifikavimo problemą.

## 2. Amžiaus kategorijų požymius klasifikuojančių tinklų analizė

Prieš atliekant amžiaus cenzo vertinimą žaidimo turinio požymiai yra nežinomi. Todėl reikia analizuoti, ar iš turimų duomenų galima išskirti reikalingus požymius ir pagal juos nustatyti žaidimo amžiaus kategoriją.

Skirtingi požymiai reikalauja skirtingų duomenų tipų. Pirmas pagal įtaką požymis – smurtas – gali būti atpažįstamas tiek vaizdiniuose, tiek garsiniuose duomenyse. Tačiau antras pagal įtaką požymis – stipriai vulgari kalba – akivaizdžiai turėtų būti ieškomas garsiniuose duomenyse.

Apie žaidimą dažniausiai prieinami duomenys yra žaidimo eigoje užfiksuoti statiniai ir dinaminiai vaizdai su garsu, taip pat reklaminė video medžiaga su vaizdu ir garsu, tekstiniai aprašai.

Žinant kokie požymiai aktualūs amžiaus kategorijai nustatyti galima analizuoti egzistuojančia literatūrą apie tokių požymių išrinkimą. Minėtų požymių išrinkimas yra aktualus ne tik kompiuteriniams žaidimams, todėl galima analizuoti literatūrą, kaip panašias problemas sprendžia kitose medijos srityse: filmuose, muzikoje, tekstuose ir t. t.

Sekančiuose skyreliuose apžvelgiama kaip panašias klasifikavimo problemas sprendžia analizuojant skirtingo tipo duomenis: vaizdą, garsą ir tekstą. Kadangi skirtingiems požymiams nustatyti galimai reikia skirtingų duomenų tipų, taip pat apžvelgiamos daugiaryšės (angl. *multimodal*) architektūros, kur tinklas lygiagrečiai priima skirtingus duomenų tipus ir išskiria duomenų požymius iš kiekvieno atskirai. Naudojant tokią architektūrą būtų galima sujungti vaizdą ir garsą, ar net tekstinį žaidimo aprašymą.

### 2.1. Vaizdiniai duomenys

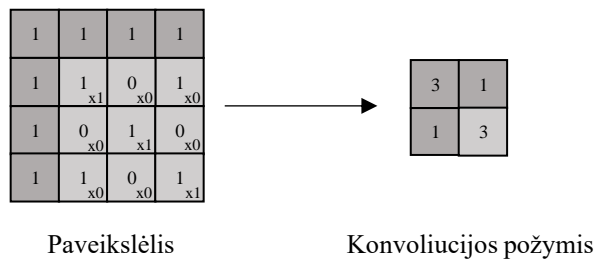
Smurto paieška vaizdiniuose duomenyse yra žinoma ir sprendžiama problema. Skirtingi modeliai šį uždavinį sprendžia skirtingai, t. y. naudoja skirtingas tinklo architektūras. Tačiau daugelis modelių vaizdinės medžiagos požymiams išskirti naudoja bendrą komponentą – konvoliucinius neuroninius tinklus [SL17, DFZ+14].

#### 2.1.1. Konvoliuciniai neuroniniai tinklai

Vaizdinės medžiagos klasifikavimui sėkmingai naudojami konvoliuciniai neuroniniai tinklai (angl. *convolutional neural networks*, toliau KNT). KNT yra gilusis neuroninis tinklas, kuriam būdingas automatinis aktualių požymių (angl. *features*) išrinkimas be žmogaus įsikišimo. Ši savybė leidžia mokymosi metu apdoroti didelį kiekį duomenų ir išrinkti svarbiausius požymius, kur tai būtų neįmanoma atlikti rankiniu būdu [LBH15, STT17].

KNT apdoroja N dimensijų matricos pavidalo įvestis, kur kiekvienas elementas atitinka vaizdo taško ryškumą. Statinių vaizdų atveju tai gali būti dvejų dimensijų matrica nespalvotiems vaizdams, arba trijų dimensijų matrica spalvotiems, kuri susideda iš trijų spalvos kanalų ir dvimačių matricių [LBH15].

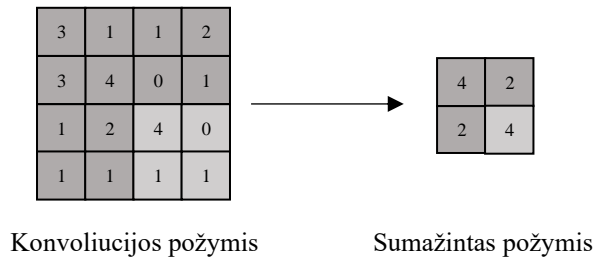
Įvesties matricos apdorojamos keliuose neuronų sluoksniuose. KNT sluoksniai, kitaip nei standartiniai GNT susideda ne iš paprastų neuronų, bet iš filtrų, dar vadinamus branduoliais (angl. *kernels*). Filtrai kaip įvestį gauna vaizdo medžiagą matricos pavidalu ir su ja, slenkant per kiekvieną elementą, atlieka konvoliucijos operaciją. Kiekvieno filtro išvestyje gaunamas vaizdas, kuriame yra išryškintas požymis, kurio atitinkamas filtras ieško. Konvoliucijos operacija pavaizduota 2 pav. [LBH15].



2 pav. Dvimatė konvoliucijos operacija

Klasifikavimo uždavinio atveju išskirti požymiai perduodami į pilnai jungių neuronų sluoksnius, kurių išvestis yra spėjamų klasių tikimybės.

Požymių matricos padeda klasifikuoti vaizdus, tačiau jų dydis kelia skaičiavimo resursų problemą. Dirbant su didesniais paveikslėliais gaunama požymių matrica yra taip pat didelė. Remiantis idėja, kad požymiai lokaliaje vaizdo erdvėje yra maždaug vienodi, matricą galima sumažinti. Mažinimas vykdomas pasitelkus mažinimo matricą kartu su kokia nors funkcija. Tokį mažinimą atlieka mažinimo sluoksnis (angl. *pooling layer*). 3 pav. pavaizduota maksimumo mažinimo operacija (angl. *maximum pool operation*), kurios metu iš slenkamos matricos parenkama didžiausia elemento reikšmė ir ji perkeliama į naują matricą. Gauta matrica yra mažesnė, tačiau išsaugo ryškiausius požymius [STT17].



*3 pav. Maksimumo mažinimo operacija*

Egzistuoja galimybė naudoti iš anksto apmokytus KNT. Tokie KNT dažnai pasižymi labai geru požymių išrinkimu. Kadangi tinklai dažniausiai apmokomi su labai didelėmis duomenų aibėmis, tai jų požymių išrinkimo savybės yra labai bendrinės ir tinka spręsti įvairius uždavinius su įvairiomis duomenų aibėmis. Viena populiariausių tokių vaizdinių didelių duomenų aibių yra *ImageNet*. Šioje aibėje yra daugiau nei 15 mln. aukštos kokybės suklasifikuotų nuotraukų, kurios skirstomos į daugiau nei 22 tūkst. kategorijų. Naudojant šią aibę kasmet yra rengiamos vaizdus klasifikuojančių algoritmų varžybos. Šios varžybos tapo svarbiu vaizdus klasifikuojančių algoritmų lyginimo būdu. 2012 metais šiose varžybose pirmą kartą buvo panaudotas GNT pavadinimu *AlexNet*, kuris klaidos tikimybę sumažino ~10 % nuo sekančio geriausio klasikinio algoritmo, kuriame vaizdų požymiai išrenkami rankiniu būdu. Tolimesnėse varžybose buvo pasiūlyta vis daugiau GNT, kurie rodė vis geresnius rezultatus. Daugelis šių tinklų gali būti panaudojami sprendžiant skirtingus uždavinius.

### **2.1.2. Video medžiagos įvestis**

Dinaminiai vaizdai (video medžiaga) gali turėti privalumų lyginant su statiniais vaizdais. Vaizdų keitimasis laike teikia papildomą informaciją, ypač nagrinėjant judesius.

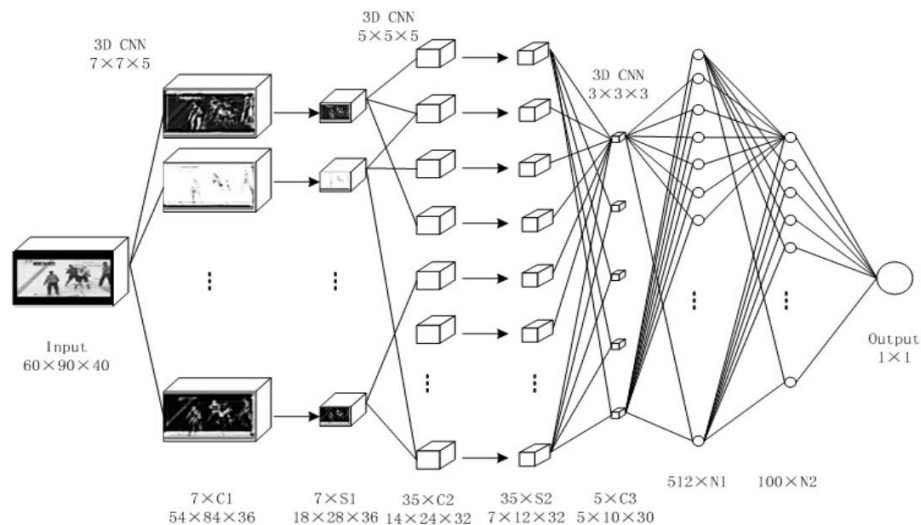
Smurtinių veikslių dažniausiai ieškoma ne statiniuose vaizduose, bet video medžiagoje. Iš dalies tai kyla iš pačios problemos, kai smurto ieškoma filmuose, internetiniuose vaizdo klipuose, ar apsaugos kamerų įrašuose. Iš kitos pusės smurtas dažniausiai pasireiškia judesiu, kuriam klasifikuoti padeda video medžiagoje esanti objektų judėjimo informacija [DZW+15].

Video medžiagą analizuojančių tinklų architektūra skiriasi nuo statinius vaizdus analizuojančių tinklų. Būdu, kaip analizuoti video medžiagą taip pat yra įvairių [DFZ+14, SL17].

Darbo [DFZ+14] autoriai sėkmingai panaudojo 3-jų dimensijų KNT smurtui aptikti ledo ritulio varžybų vaizdinėje medžiagoje ir pasiekė 91 % tikslumą.

3-jų dimensijų KNT vietoje spalvos kanalo yra laikomi video kadrai. Standartiškai KNT filtrai išrenka požymius iš statinės vaizdinės medžiagos. Išplėtus filtrus iki 3-jų dimensijų, gausime filtrus,

kurie požymius išrenka iš kelių kadru iš karto ir taip išskiria požymius apie vaizdinės medžiagos kitimą laike. Kitos tinklo savybės išlieka tokios pačios, pvz. mažinimo sluoksnis veikia taip pačiai, kaip ir statinių vaizdų atveju.



4 pav. [DFZ+14] tinklo architektūra su 3-jų dimensijų filtrais

Darbe [DFZ+14] aprašyto tinklo architektūrą susideda iš 9 sluoksnių. 4 pav. pateikta tinklo architektūros diagrama. Pirmas sluoksnis yra įvestis – trimatė 60x90x40 dydžio matrica. Antrame sluoksnyje pritaikomas trimatis 7x7x5 dydžio filtras. Paskutinė koordinatė nurodo, kad filtras išskiria 5 kadru ilgio požymius. Toliau seka mažinimo sluoksnis, KNT sluoksnis, mažinimo sluoksnis, dar vienas KNT sluoksnis ir 2 pilnai jungūs sluoksniai klasifikuoti KNT išrinktus požymius. Išvestis yra viena reikšmė parodanti ar išanalizuotoje video ištraukoje yra smurtinių veiksmų.

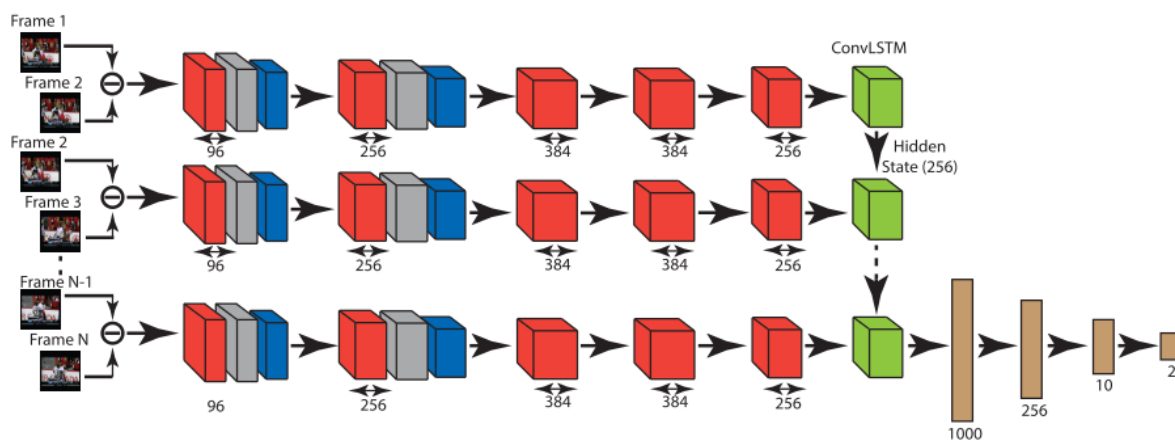
Įvesties trečiosios dimensijos dydis apriboja, kokio ilgio (kiek kadru) video medžiagą tinklas gali apdoroti. Kadangi atmintis ir skaičiavimo resursai yra apriboti, tai toks 3-jų dimensijų KNT negali apdoroti ilgos video medžiagos. [DFZ+14] atveju įvestis buvo 60x90x40, t. y. 60 taškų aukščio, 90 taškų pločio ir 40 kadru trukmės video medžiaga. Kai klasifikavimo tikslas yra nustatyti ar analizuojamas video fragmentas vaizduoja smurtą ar ne, tai tiek kadru užtenka. Tačiau žaidimo amžiaus kategorijos klasifikavimo uždavinyje įvertinti reikia ne žaidimo ištrauką, o visą žaidimą. Kadangi tikimybė, kad 40 atsitiktinai pasirinktų kadru ištrauka turės visus reikalingus amžiaus cenzo požymius yra maža, tai ir tikslaus sprendimo apie žaidimo amžiaus kategoriją priimti iš 40 kadru nepavyktų.

[SL17] panaudojo standartinius KNT kartu su konvoliucine ilga trumpalaikė atmintimi (angl. *convolutional long short-term memory*, toliau KonvLSTM). Šis tinklas ledo ritulio smurto aptikimo duomenų aibėje pasiekė žymiai geresnę 97 % tikslumą, nei 3-jų dimensijų KNT. Kadru požymiai buvo individualiai išrenkami naudojant KNT ir tada agreguojami naudojant KonvLSTM.

Kadru požymių išrinkimui buvo naudojamas iš anksto apmokytas KNT pavadinimu AlexNet. Šis tinklas apmokytas naudojantis didele vaizdų aibe ImageNet ir pasižymi labai geru bendrinių požymių išrinkimu. Naudojant *AlexNet*, vietoje iš anksto neapmokyto KNT, smurto klasifikavimo tikslumas ledo ritulio aibėje padidėjo ~2 %.

Požymius agreguojantis KonvLSTM yra rekurentinio neuroninio tinklo su ilga trumpalaikė atmintimi (angl. *long short-term memory*, toliau LSTM) variantas. LSTM duomenis priima vykdant kelis žingsnius, kur kiekviename žingsnyje į įvestį taip pat ateina prieš tai buvusio žingsnio LSTM išėjimas. Tokiu būdu LSTM išsaugo informaciją tarp skirtingų žingsnių. Priešingai nei LSTM, KonvLSTM duomenis išlaiko matricos struktūroje. Tokiu būdu užfiksuojami kadru požymių kitimai ne tik laike, bet ir erdvėje.

Taip pat [SL17] pasiūlė pakeisti kadru įvestį iš individualių kadru į šalia esančių kadru skirtumą. Tokiu būdu įvesties duomenyse iš karto yra užfiksuojamas vaizdinis pokytis. Tai reiškia, kad tinklas privalomai atsižvelgs į vaizde vykstančius pokyčius. Šis pakitimas smurto aptikimo tikslumą ledo ritulio duomenų aibėje padidino ~1 %.



5 pav. [SL17] tinklo architektūra priimanti kadru skirtumus ir agreguojanti jų požymius KonvLSTM sluoksnyje

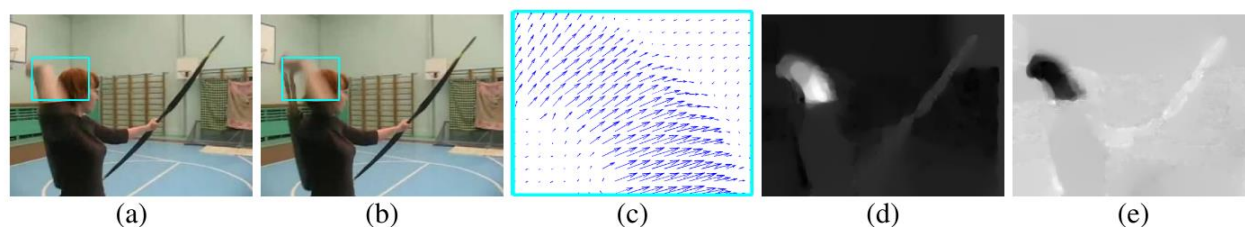
Darbe [SL17] aprašyto tinklo architektūra pateikta 5 pav. Tinklas susideda iš 11 sluoksnių. Prieš pateikiant duomenis į įvestį atliekami tam tikri skaičiavimai: atrenkama N vienodai nutolusių kadru,



kadrai sumažinami į 256 taškų pločio ir 256 taškų aukščio matricas ir paskaičiuojami skirtumai tarp šalia esančių kadru. Gauti duomenys pateikiami į tinklą. Paveikslėlyje raudonai pavaizduoti KNT sluoksniai, pilkai – normalizavimo sluoksniai, mėlynai – mažinimo sluoksniai. Paminėti sluoksniai atsakingi už kadru požymių išrinkimą. Išrinkti požymiai toliau perduodami į žalią – KonvLSTM sluoksnį, kur požymiai agreguojami priimant kiekvieną įvestį. Atlikus visus įvesties žingsnius KonvLSTM suagreguoti požymiai laike pateikiami į pilnai jungius sluoksnis, kur vykdomas klasifikavimas.

KonvLSTM naudojimas vietoje 3-jų dimensijų KNT techniškai leidžia įvesti bet kokio ilgio video medžiagą. Tačiau ilgesnė video medžiaga gali sukelti kitų problemų: našumo sumažėjimas (reikia apdoroti daugiau duomenų), neteisinga scenų pasikeitimo interpretacija (staiga pasikeitęs kadras turės daug skirtumų su prieš tai buvusiu).

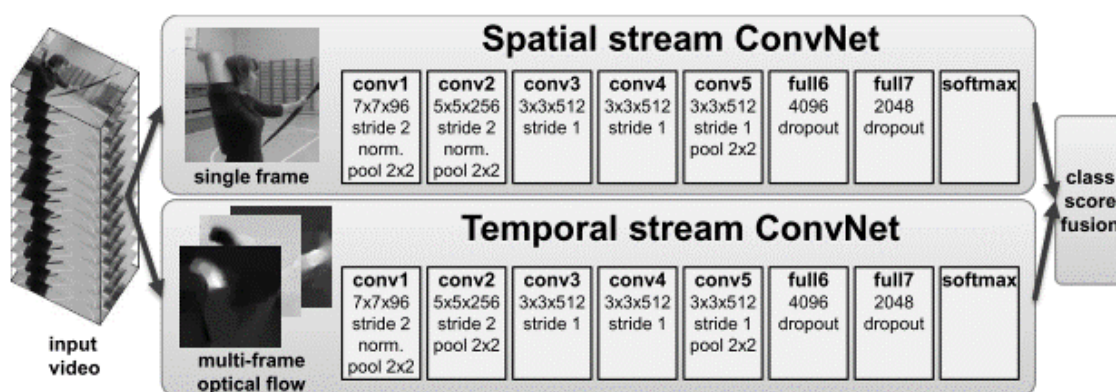
Darbai su ilga video medžiaga [NHV+15] pasiūlė išretintų kadru komponavimą kartu su optiniais srautais (angl. *optical flow*) pagal [SZ14] pavyzdį. Optiniai srautai, panašiai kaip [SL17] skaičiuojamas skirtumus tarp šalia esančių kadru, užfiksuoja informaciją, kaip objektai juda erdvėje bėgant laikui. Skirtumas toks, kad [SL17] atveju lyginami tik 2-jų šalia esančių kadru skirtumas, o naudojant optinius srautus galima užfiksuoti aiškesnius ir ilgesnius judesius. [NHV+15] vienu kartu galėjo apdoroti iki 2 minučių ilgio video medžiagą (120 kadru). Optinių srautų pavyzdys pateiktas 6 pav. Paveikslėlyje vaizduojami apskaičiuoti optiniai srautai tarp dviejų kadru pažymėtų (a) ir (b). Paveikslėlyje (c) matomi vaizdo taškų poslinkio vektoriai. Paveikslėliuose (d) ir (e) yra užkoduotos vektorių horizontalios ir vertikalios reikšmės. Optinio srauto vaizde aiškiai matome užfiksuotą rankos ir lanko judėjimo informaciją. Šiuos duomenis, kaip ir paprasta vaizdinę medžiagą, sėkmingai apdoroja KNT. Optinių srautų trūkumas – reikalingi papildomi skaičiavimai sukurti judesio vektorius, prieš juos pateikiant į tinklą.



6 pav. Optiniai laukai tarp dviejų kadru [SZ14]

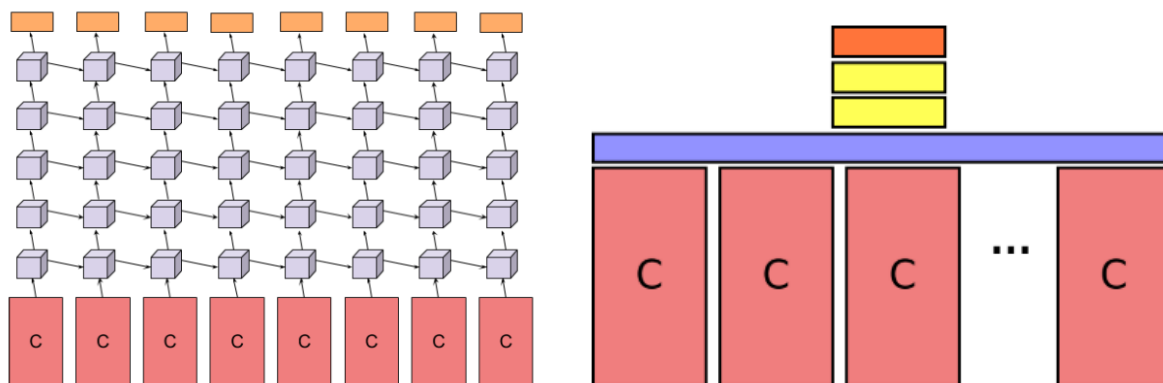
Siekiant panaudoti ir vaizdinę ir judesio informaciją naudojama dviejų srovių KNT (angl. *two-stream convolutional network*) architektūra pavaizduota 7 pav. Ši architektūra susideda iš dviejų paprastų KNT, tačiau vienas priima vaizdinę kadro medžiagą, kitas optinių srautų vaizdą. Abu tinklai išrenka iš pateiktų duomenų požymius ir pilnai jungiame sluoksnyje padaro nepriklausomus klasifikavimo spėjimus. Rezultate abu spėjimai apjungiami, pasiūlytu atveju tiesiog apskaičiuojant vidurkį.

Darbo [DZW+15] autoriai palygino klasifikatorių priimančių pavienius kadrus ir dviejų srovių klasifikatorių optiniais srautais ir LSTM agregavimo sluoksniu. Dviejų srovių architektūros tinklas pagerino rezultatą 10 %.



7 pav. Dviejų srovių tinklas [SZ14]

Darbo [NHV+15] autoriai palygino du skirtingus būdus agreguoti vaizdų ir optinių srautų požymius laike: LSTM sluoksnį ir mažinimo sluoksnį, kuris priima daug lygiagrečių KNT, apdorojančių kiekvieną kadrą atskirai. Abi architektūros pavaizduotos 8 pav. Kairėje pusėje matoma LSTM architektūra, kur raudonai pažymėti blokai simbolizuoja požymius išskiriančius KNT sluoksnius, mėlsvai pažymėti 5 LSTM sluoksniai ir oranžine spalva pažymėti išvesties funkcija. Dešinėje pusėje matoma lygiagrečių KNT sluoksnių architektūra, kur raudonai taip pat pažymėta KNT sluoksniai, kiekvienas priimančys iš eilės einančius kadrus. Tada visų kadrų požymiai apjungiami mėlynai pažymėtame mažinimo sluoksnyje. Suagreguoti požymiai perduodami į geltonai pažymėtus pilnai jungius sluoksnius ir galiausiai perduodami išvesties funkcijai. Abejais atvejais vaizdų ir optinių laukų klasifikacija vyko atskirai ir rezultatai buvo apjungiami taip pačiai kaip [SZ14]. Abi architektūros rodė gerus rezultatus, tačiau LSTM buvo šiek tiek tikslesnė: nagrinėjant UCF-101 veiksmų aibę LSTM architektūra pasiekė 88,6 % tikslumą, o lygiagrečių KNT – 88,2 %.



8 pav. [NHV+15] ilgos-trumpos atminties ir lygiagrečių KNT architektūros

Taigi, aptartos architektūros parodo, kad analizuojant video medžiagą galima nustatyti smurtinius ar kitokius veiksmus dideliu tikslumu. Kadangi smurtas yra vienas iš didžiausių žaidimų amžiaus kategorijos nustatymui reikalingų požymių, galima manyti, kad aptarti tinklai turėtų rodyti neblogus rezultatus ir klasifikuojant žaidimų amžiaus kategorijas.

## 2.2. Garsiniai duomenys

Žaidimai turi labai daug skirtingų garsų, kurie potencialiai gali teikti daug svarbios informacijos, pvz., veikėjų dialogai, garso efektai tokie, kaip šūviai, sprogimai, foniniai garsai nusakantys vietą, muzika nusakanti būseną.

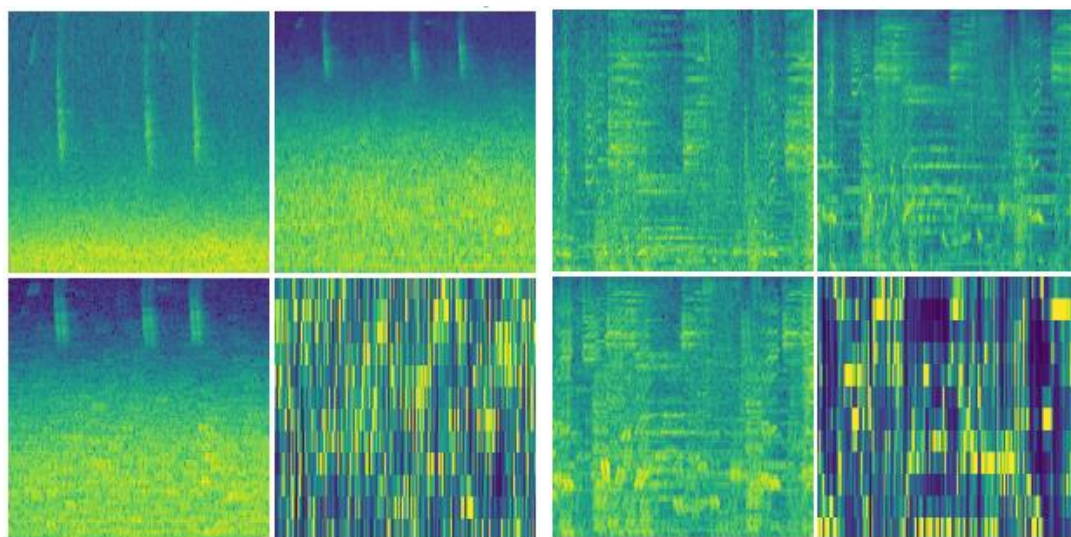
Garsas padengia kelis amžiaus kategorijų įtakojančius požymius. Kai kurie smurto požymiai pasireiškia ne tik vaizdu bet ir garsu: riksmi, sprogimai, šūviai. Kai kurie smurto požymiai garsu gali pasireikšti aiškiau, nei vaizdu, pvz. šūviai aiškiai girdisi, tuo tarpu vizualiai gali matytis labai trumpai arba visai nesimatyti. Taip pat garsas yra pagrindinis būdas nustatyti vulgarią kalbą.

Darbo [SSI+13] autoriai nustatė, kad bandant atpažinti smurtą garsiniai duomenys gali išgauti didesnę tikslumą, nei vaizdiniai duomenys. Buvo bandoma klasifikuoti smurtą ištraukose iš filmų. Pagal gautus rezultatus matoma, kad garsiniai duomenys pateikė 14 % didesnę tikslumą nei pavieniai vaizdiniai kadrai.

Kito darbo [ACG+19] autoriai pateikė tinklą, kuris naudodamasis tik garsine informacija nustato žaidimo žanrą iš 6 galimų 66,9 % tikslumu. Tarp žaidimo žanro ir amžiaus kategorijos gali egzistuoti ryšys. Pavyzdžiui, tikėtina, kad kariniai ar siaubo žaidimai turės griežtesnę amžiaus kategoriją nei simuliacijos ar sporto žaidimai, kadangi kariniams ir siaubo žaidimams dažniau būdingas smurtas ir kiti panašūs požymiai. Todėl gaunama informacija, kurią tinklas naudoja nustatydamas žaidimo žanrą, potencialiai gali būti naudinga ir nustatant amžiaus kategoriją.

Tinklas garsą gali priimti kaip vaizdinę informaciją spektrogramos formatu. Spektrograma vaizduoja kaip garso dažnių intensyvumas kinta laike. Požymiai iš spektrogramos išrenkami naudojant KNT. Priešingai nei su vaizdu, KNT tinka analizuoti garso pokyčius laike, kadangi spektrograma koduoja garso dažnių stiprumą laike. 9 pav. pavaizduotos dviejų skirtingų žanrų žaidimų garso ištraukų spektrogramos. Galime matyti akivaizdžius vizualius skirtumus. Taigi KNT iš tokių vaizdų išrenka požymius, pagal kuriuos galima atlikti klasifikavimą.

Autoriai pastebėjo, kad iš anksto apmokyti KNT (pvz., *AlexNet*) sėkmingai išrenka reikalingus požymius iš spektrogramos vaizdų, nors buvo mokyti su paprastomis nuotraukomis. Tai parodo, kad KNT apmokyti ant didelių duomenų aibių išmoksta išskirti labai bendrus požymius, kurie naudingi sprendžiant įvairius klasifikavimo uždavinius. Geriausi rezultatai gauti naudojantis *GoogLeNet* ir *AlexNet* tinklus.



(a) Sporto žaidimas kriketas

(b) Muštynių žaidimas

9 pav. Žaidimo garso ištraukų spektrogramos [ACG+19]

Darbo [CPH+17] autoriai pateikė tinklą, kuris iš garsų spektrogramų nustato įvairių garsų kilmę, pvz. šaunamojo ginklo šūvį atpažįsta 73 % tikslumu. Šiame tinkle, kaip ir anksčiau minėtuose tinkluose, iš garso spektrogramų požymiai išrenkami KNT sluoksniuose. Tačiau šis tinklas priima kelis spektrogramos kadrus ir užfiksuoja garso pokyčius laike LSTM sluoksniuose. Kadangi su KNT tinklais naudojami mažinimo sluoksniai, kurie spektrogramų atveju gali iškraipyti laiko informaciją, buvo taikomi mažinimo sluoksniai tik dažnių ašiai, taip išlaikant laiko mastelį vientisą tarp skirtingų kadrų ir leidžiant LSTM tiksliau atpažinti vykstančius pokyčius laike. Suagreguoti LSTM sluoksnyje požymiai perduodami į pilnai jungtą sluoksnį, kur vyksta klasifikacija.

Darbo [HBC+18] autoriai taip pat naudojo garso spektrogramas tyrinėti žaidimuose kalbančių veikėjų kalbą. Buvo bandoma klasifikuoti skirtingų veikėjų akcentus. Autoriai panaudojo *AlexNet* architektūrą ir apmokė tinklą su dviejų akcentų duomenų aibe. Klasifikuojant britišką ir amerikietišką akcentus pavyko pasiekti 71 % tikslumą.

Tinklo gebėjimas analizuoti kalbos garsus skamba žadančiai bandant išrinkti amžiaus kategorijų požymius. Balso analizė būtų naudinga atpažinti vulgarios kalbos požymius, taip pat smurtą, baimę ir kt.

### 2.3. Tekstiniai duomenys

Daugelis žaidimų turi tekstinį aprašą, kuris gali teikti naudingą informaciją. Teksto formulavimas ir pasirenkami žodžiai gali indikuoti kokio amžiaus auditorijai žaidimas skirtas. Taip pat tekste galima pastebėti žodžių susijusių su smurtu, vulgaria kalba, tabako, alkoholio ar narkotikų vartojimu.

Teksto apdorojimui, kaip ir kitiems duomenų tipams, yra skirtingų būdų. Vienas iš geriausių rezultatus duodančių būdų yra paversti tekste esančius žodžius į N-matės erdvės vektorius, taip, kad vektoriai iš žodžių pasirodančių panašiam kontekste būtų ne toli vienas nuo kito. Toks žodžių paskirstymas padeda apie žodį sutalpinti daug informacijos į mažą kiekį duomenų [MSC+13].

Darbo [MSC+13] autoriai pasiūlė efektyvų žodžių į vektorius konvertavimo metodą pavadinimu *word2vec*. Šis metodas remiasi neuroniniu tinklu, kuris yra apmokomas surasti dažnai kartu vartojamus žodžius pagal įvestą žodį. Autoriai išbandė tinklą apmokyti su 30 milijardų žodžių teksto ir gavo labai gerus rezultatus. Atliekant operacijas su žodžių vektoriais buvo gaunami prasmingi rezultatai, pvz., iš žodžio „Madridas“ vektoriaus atėmus žodžio „Ispanija“ vektorių ir pridėjus žodžio „Prancūzija“ vektorių gauname vektorių, kurio artimiausias žodis yra „Paryžius“. Šis pavyzdys parodo, kaip vektoriaus koordinatėse yra užkoduota pakankamai naudinga informacija nustatyti, kad žodis reiškia atitinkamos šalies sostinę. Kitas pavyzdys prie žodžio „Vokietija“ vektoriaus pridėjus žodžio „sostinė“ vektorių gauname vektorių artimą žodžiui „Berlynas“.

Kito darbo [MM15] autoriai sprendė muzikos amžiaus kategorijos klasifikavimo uždavinį. Sprendimas nagrinėjo albumuose esančių dainų tekstus ir bandė priskirti albumui vieną iš sekančių amžiaus kategorijų: kūdikiams (2–3 m.), ikimokyklinukams (4–5 m.), mokiniams (6–8 m.), mokiniams (9–11 m.), jauniems paaugliams (12–14 m.), paaugliams (15–17 m.), suaugusiems (>17 m.). Žodžių požymiai išrenkami naudojantis iš anksto apmokytu žodžius į vektorius verčiančiu

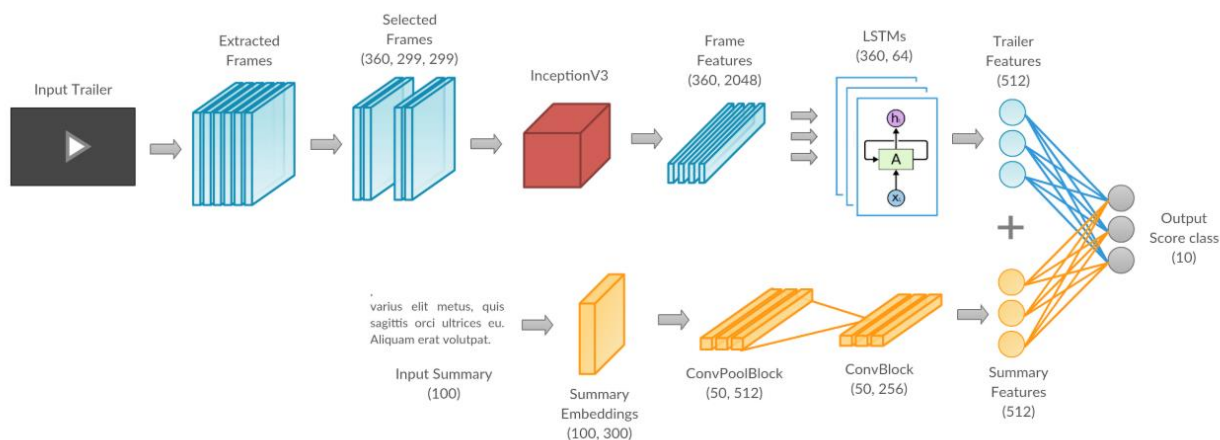


įrankiu *GloVe*. Taip pat prie žodžių požymių buvo pridėtos įvairios žodžių savybės (pvz., būdingas amžius kada ieškomo žodžio išmokstama) iš MRC psicholingvistinės duomenų bazės. Pagal išrinktus požymius klasifikavimas vyko naudojant *M5P Weka* klasifikatorių. Pasiūlytas sprendimas pasiekė ~71 % tikslumą.

## 2.4. Daugiarūšės tinklų architektūros

Visi nagrinėti duomenų tipai atrodo naudingi ir potencialiai teikiantys informaciją apie amžiaus kategorijos požymius. Todėl svarbu išbandyti kaip vertinimo tikslumą įtakoja klasifikavimas pagal kelis duomenų tipus iš karto, palyginus su klasifikavimu pagal vieną duomenų tipą. Neuroniniai tinklai, kurie gali priimti skirtingus duomenų tipus vienu metu, yra vadinami daugiaryšiais neuroniniais tinklais (angl. *multimodal neural networks*).

Darbo [BBR+18] autoriai pateikė žaidimo reitingą spėjantį daugiaryšį neuroninį tinklą. Tinklas kaip įvestį priima žaidimo video anonsą ir tekstinį aprašą. Kaip rezultatą bando atspėti žaidimo žaidėjų ir kritikų reitingą šimtabalėje sistemoje. Klasifikuojama 10 klasių, kur kiekviena klasė atitinka vertinimo intervalą šimtabalėje sistemoje (pvz., 0–10, 11–20 ir t. t.). Tinklas atspėja kritikų ir žaidėjų reitingo intervalą 70,5 % tikslumu. Autoriai parodė, kad teksto aprašų prijungimas prie klasifikatoriaus tikslumą padidino 4 %.



10 pav. Daugiaryšis tinklas numatantis kritikų ir žaidėjų vertinimą [BBR+18]

Tinklas susideda iš dviejų dalių (žr. 10 pav.). Mėlynai pažymėta dalis apdoroja vaizdinę medžiagą. Kaip įvestį priima iš video anonso algoritmiškai išrinktus kadrus. Iš kadru išrenkami požymiai naudojantis iš anksto apmokytu tinklu *InceptionV3*. Išrinkti požymiai agreguojami LSTM

sluoksnyje. Apdoroti kadru požymiai perduodami į pirmus 512 pilnai jungaus sluoksniu perceptronus.

Antroji tinklo dalis (pažymėta geltonai) apdoroja tekstinę medžiagą. Tinklas priima į vektorius paverstus 100 žodžių. Iš pradžių visi žodžiai yra pakeičiami skaičiumi sudarant bendrą žodyną iš visų turimų mokymosi duomenų. Tada toks skaičių rinkinys yra sutrumpinamas iki 100 skaičių, arba papildomas nuliais, jeigu nesiekia 100. Šiuos apdorotus duomenis galima pateikti tinklui. Pirmame sluoksnyje vyksta teksto įterpimas (angl. *embedding*), kur skaičiai yra paverčiami į 300 elementų ilgio vektorius pagal tai kiek žodžiai yra prasmingi tarpusavyje. Iš šios matricos išrenkami požymiai naudojantis paprastais KNT ir mažinimo sluoksniais. Požymiai perduodami į tą patį pilnai jungų sluoksnį, kuriame saugomi ir vaizdinės medžiagos požymiai, tik šį kartą užpildomi paskutiniai 512 perceptronų. Tokiu būdu pilnai jungus sluoksniu yra užpildomas tekstiniais žaidimo požymiais ir video anonso požymiais. Naudojantis visais požymiais paskaičiuojama tinklo išeitis (spėjamas reitingas).

### 3. Duomenų aibės sukūrimas

Kompiuterinių žaidimų amžiaus cenzo klasifikavimo uždaviniui spręsti reikalinga duomenų aibė su visais prieš tai minėtais duomenų tipais. Duomenų aibės elementai turi būti žaidimai, jų ištraukų video įrašai su garsu, tekstiniai aprašai ir amžiaus cenzo kategorijų įvertinimai. Kadangi duomenų aibė, turinti visus minėtus reikalavimus, nebuvo rasta, tai reikėjo sukurti naują duomenų aibę.

Sukurtą duomenų aibę sudaro kompiuterinių žaidimų multimedija ir metaduomenys. Multimediją sudaro spalvoti judantys ir statiniai vaizdai, garsai ir tekstinis žaidimo aprašas. Metaduomenis sudaro žaidimo pavadinimas, originalus amžiaus cenzo įvertis pagal ESRB arba PEGI, viena iš trijų galimų amžiaus cenzo klasių, vaizdo įrašo nuoroda ir vaizdo įrašo kokybė.

Duomenų aibėje yra 1116 žaidimų. Aibė padalinta į mokymosi ir testavimo poaibius santykiu 80 % ir 20 %. Siekiant išlaikyti klasių balansą, kiekvieną poaibio duomenų tipo ir klasės grupę sudaro po vienodas skaičius įrašų. Detali duomenų aibės sudėtis pateikta 3 lentelėje. Duomenų aibės dydis diske – 121 GB.

3 lentelė. Duomenų aibės sudėtis

Aibė Klasė Duomenys	Mokymosi			Testavimo		
	Vaikams	Paaugliams	Suaugusiems	Vaikams	Paaugliams	Suaugusiems
Žaidimai	297	297	297	75	75	75
Tekstiniai aprašai	297	297	297	75	75	75
Vaizdo įrašų ištraukos	2 970	2 970	2 970	750	750	750
Vaizdo įrašų kadrai	53 460	53 460	53 460	13 500	13 500	13 500
Optinių srautų kadrai	53 460	53 460	53 460	13 500	13 500	13 500

#### 3.1. Žaidimų sąrašo sudarymas

Žaidimų sąrašas sudarytas naudojantis internetine kompiuterinių žaidimų duomenų baze (angl. *internet game database*, toliau IGDB). Duomenų bazė pateikia tinklinę sąsają, priimančią *Apicalypse* formato užklausas ir grąžina rezultatus *JSON* formatu. Sąrašas sudarytas pagal užklausas su tokiais kriterijais:

- Žaidimas įvertintas ESRB arba PEGI vertinimo skalėje.
- Žaidimas turi tekstinį aprašą.



- Žaidimas įvertintas bent 2 kartus (taip siekiama atrinkti labiau žinomus žaidimus ir padidinti tikimybę rasti žaidimo vaizdo įrašą).

IGDB saugo žaidimų meta duomenis ir statinius vaizdus (angl. *screenshots*), tačiau neturi vaizdo įrašų. Todėl, siekiant duomenų kokybės vientisumo tarp statinių ir dinaminių vaizdų, multimedijos duomenų buvo ieškoma kitur, o iš šios duomenų bazės buvo atrinktas tik žaidimų sąrašas su amžiaus cenzo įvertinimais ir tekstiniais aprašais. Šie duomenys saugomi kiekvieno žaidimo kataloge, faile pavadinimu *meta.json*. Meta duomenų pavyzdys pateiktas 11 pav.

```
{
  "Name": "Minecraft Dungeons",
  "Summary": "\u0022Brave the dungeons alone, or team up with friends! ...",
  "Rating": 1,
  "OriginalRating": "E10",
  "OriginalName": "Minecraft Dungeons",
  "VideoUrl": "https://www.youtube.com/watch?v=NMYdvrt28E",
  "VideoQuality": "360p"
}
```

11 pav. Meta duomenų failo pavyzdys.

Siekiant padidinti duomenų aibės imtį ir pagerinti klasifikatoriaus mokymosi galimybes, PEGI ir ESRB vertinti žaidimai buvo apjungti į vieną aibę su suvienodintomis ir supaprastintomis vertinimo kategorijomis. Kadangi PEGI ir ESRB amžiaus kategorijos naudoja skirtingus metų intervalus, siekiant apjungti šių sistemų vertinimus reikia apjungti ir tam tikras kategorijas. Visi žaidimai buvo suklasifikuoti į tris amžiaus kategorijas: 1 – „vaikams“, 2 – „paaugliams“ ir 3 – „suaugusiems“. Tikslus apjungimas pavaizduotas 4 lentelėje. Šis apjungimas paremtas PEGI požymių panašumu tarp apjungtų kategorijų ir minimaliu persidengimu tarp kategorijų ESRB atžvilgiu.

4 lentelė. PEGI ir ESRB amžiaus grupių apjungimas į tris klases

ESRB	PEGI	Klasė (amžius)
Ankstyvai vaikystei (angl. <i>early childhood</i> ) 3+	3+	Vaikams [3; 12)
Visiems (angl. <i>everyone</i> ) 6+	7+	
Visiems 10+	12+	
Paaugliams (angl. <i>teen</i> ) 13+	16+	Paaugliams [12;18)
Subrendusiems (angl. <i>mature</i> ) 17+	18+	
Suaugusiems (angl. <i>adults only</i> ) 18+	18+	Suaugusiems [18; +∞)

### 3.2. Vaizdinė medžiaga

Duomenų aibei sudaryti reikalingi žaidimo metu užfiksuoti vaizdo įrašai. Vaizdo įrašė turi būti tik su žaidimu susijusi vaizdo medžiaga, be žaidėjo komentarų, be vaizdo kamera užfiksuoto veido ir be reklaminių ar kitokių įterpimų.

Vaizdo įrašai surinkti iš vaizdo įrašų dalijimosi platformos *YouTube*. Vaizdo įrašų ieškota programiniu būdu. Kiekvienam žaidimui atliekant paiešką pagal šabloninį tekstą „*the game \{žaidimo pavadinimas\} \"gameplay \"no commentary\”*“. Tarp gautų rezultatų atrenkami *mp4* formato vaizdo įrašai su didžiausia kokybe iš trijų galimų: *360p*, *480p* ir *240p* ir iki 50 minučių ilgio, dėl vietos diske apribojimo. Atrinkto vaizdo įrašo nuoroda išsaugota *meta.json* faile, o pats vaizdo įrašas atsiųstas ir išsaugotas kiekvieno žaidimo kataloge pavadinimu *video.mp4*.

Surinkti vaizdo įrašai sukarpyti į trumpas iškarpas. Iškarpos surinktos naudojantis *ffmpeg* įrankiu. Pirmos 30 vaizdo įrašo sekundžių atmetamos kaip nenaudotinos, nes dažnai vaizduoja su žaidimu nesusijusius vaizdus, pvz. vaizdo įrašo autoriaus prisistatymą. Likusi vaizdo įrašo dalis padalinta į 10 lygių dalių ir iš kiekvienos dalies pradžios iškirpta po apytiksliai 3 sekundes vaizdo medžiagos. Kiekviena iškarpa išsaugota žaidimo kataloge, atskirame faile pavadinimu *clipN.mp4*, kur *N* yra iškarpos indeksas nuo 0 iki 9.

Iš vaizdo įrašų iškarpų išrinkti atskiri kadrai. Kadrai išrinkti naudojantis *ffmpeg* įrankiu ir išsaugoti atskiruose failuose pavadinimu *clipN\_XXXX.jpg*, kur *N* yra iškarpos indeksas, o *XXXX* yra kadro indeksas. Iškarpos konvertuojamos į kadrus 5 kadrų per sekundę dažniu, todėl iš kiekvienos vaizdo įrašo iškarpos gauta po 18 kadrų.

Šalia kadrų taip pat sugeneruoti tankūs optiniai srautai (angl. *dense optical flow*). Optiniai srautai sugeneruoti naudojantis *Lucas-Kanade* metodu, įgyvendintu bibliotekoje *OpenCV*. Optiniai srautai sugeneruoti taip, kad atitiktų anksčiau sugeneruotus vaizdo įrašų ištraukų kadrus ir galėtų būti naudojami kartu. Kiekvienai vaizdo įrašo ištraukai sukurta po 18 optinių srautų failų.

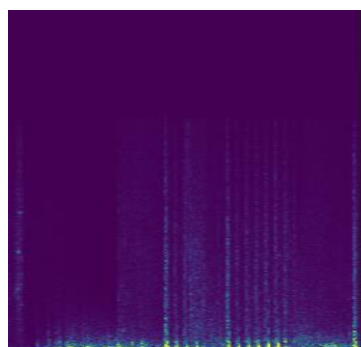
### 3.3. Garso medžiaga

Garso medžiaga išrinkta iš vaizdo įrašų iškarpų. Garso medžiaga išrinkta naudojantis *ffmpeg* įrankiu. Kiekvienos iškarpos garsas saugomas atskirame faile pavadinimu *clipN.wav*, kur *N* yra iškarpos indeksas. Visa garso medžiaga saugoma žaidimo pakatalogyje *sounds*.

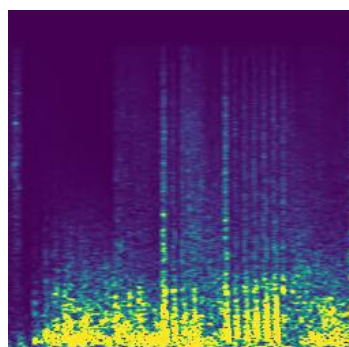
Garso įrašų iškarpos paverstos į spektrogramas. Garso failai paversti į spektrogramas naudojant *NuGet* paketą *Spectrogram*. Šis paketas taiko Furjė transformaciją, dekomponuoja nurodyto ilgio

signalų langus į atskirus dažnius ir juos pavaizduoja paveikslėlio pavidalu. Rezultatas matomas 12 pav., spektrogramoje vertikaliai paskirstytas dažnių rėžis, horizontaliai vaizduojamas pokytis laike, o spalvos ryškumas nurodo kiekvieno dažnio signalo stiprumą.

Kadangi dirbama su garsais skirtais žmogaus klausai, tai garso signalų dažnis yra apribotas. Apatinė spektrogramos dalis atitinka 0 kHz dažnius, o viršutinė – 22 kHz dažnius. Taip pat, siekiant išryškinti svarbių dažnių informaciją buvo pritaikyta *Mel* skalė. Ši skalė paskirsto garso dažnius panašiau į tai, kaip dažnius atskiria žmonių klausa, t. y. geriau atpažįstame pokyčius žemesniuose dažniuose, o aukštesniuose dažniuose eksponentiškai blogiau. 12 pav. pavaizduota to pačio tonų dažnių rėžio paveikslėliai. Matome, kad *Mel* skalėje (b) garso informacija atvaizduojama daug aiškiau, nei tiesinėje skalėje (a).



a) Garso tono dažniai tiesiškai



b) Garso tono dažniai pagal *Mel* skalę

12 pav. Garso ištraukos, iš susišaudymo žaidimo, spektrograma

### 3.4. Tekstiniai aprašai

Žaidimų tekstiniai aprašai surinkti iš IGDB. Aprašas saugomas *meta.json* faile laukelyje pavadinimu *Summary*. Aprašai duomenų aibėje saugojami nemodifikuoti. Aprašų ilgis varijuoja nuo 52 iki 9 248 simbolių, vidutinis aprašo ilgis yra 617 simbolių. Trumpiausią aprašą sudaro 8 žodžiai, ilgiausią – 1517 žodžiai, vidutiniškai aprašą sudaro – 102 žodžiai.

### 3.5. Duomenų kokybė

Duomenų kokybės užtikrinimas automatizuotas tik iš dalies. Duomenų aibę renkanti programa atmetė žaidimus, kuriuose nėra tekstinių aprašų ar atliekant vaizdo įrašų paiešką negrąžino tinkamų rezultatų.

Didžioji dalis duomenų kokybės užtikrinimo buvo atlikta rankiniu būdu. Rankinis duomenų kokybės užtikrinimas buvo vykdomas peržiūrint vaizdo įrašus ir, pagal poreikį, skaitant meta

duomenų failą. Žaidimai, turintys neatitikimų ar nekokybiškų duomenų, buvo panaikinti iš duomenų aibės ir pagal galimybes jų vietą užimdavo kitas žaidimas

Peržiūros metu panaikinti žaidimai, kurie turėjo:

1. Sugadintą vaizdo įrašą (akivaizdūs vaizdo įrašo sutrikimai, užstrigęs vaizdas ir t. t).
2. Prastą vaizdo įrašo kokybę (pvz., filmuota su rankine kamera, o ne ekraną filmuojančia programa).
3. Ne pilno dydžio vaizdo įrašą.
4. Vaizdo įrašo garso takelyje įrašytus žaidėjo komentarus.
5. Kito (ne vertinamo) žaidimo vaizdo įrašą.

### **3.6. Apibendrinimas**

Duomenų aibė sudaryta iš kelių tipų duomenų ir yra paruošta amžiaus cenzo uždaviniui spręsti, tačiau galėtų būti naudojama ir kitokiems su kompiuteriniais žaidimais susijusiais uždaviniais. Pagal poreikį, aibė lengvai plečiama. Meta duomenų failas yra *JSON* formato ir tinkamas lengvai išplėsti papildomoms klasėms. Multimedijos duomenys saugomi atskiruose aplankuose, todėl pridėti naujus duomenis taip pat nesunku.

## 4. Neuroninių tinklų kūrimas ir mokymas

Šiame skyriuje ieškoma tikslios strategijos dirbti su kiekvienu duomenų aibės duomenų tipu. Radus pranašesnę strategiją, ji bus naudojama vėlesniame modelių komponavimo žingsnyje.

Visi sekantys neuroniniai tinklai buvo kuriami ir mokomi vienoje sistemoje. Neuroninių tinklų programos rašytos programavimo kalba *Python* (versija 3.7.9). Neuroninių tinklų kūrimui ir apmokymui naudota biblioteka *Keras* (versija 2.4.0) su *TensorFlow* (versija 2.4.0) įgyvendinimu. Skaičiavimai vykdomi naudojantis *Nvidia GeForce GTX 1070* grafiniu procesoriumi su 8 GB atminties, taip pat *Intel i7-4770* centriniu procesoriumi ir 16 GB darbinės atminties.

### 4.1. Darbas su statine vaizdine medžiaga

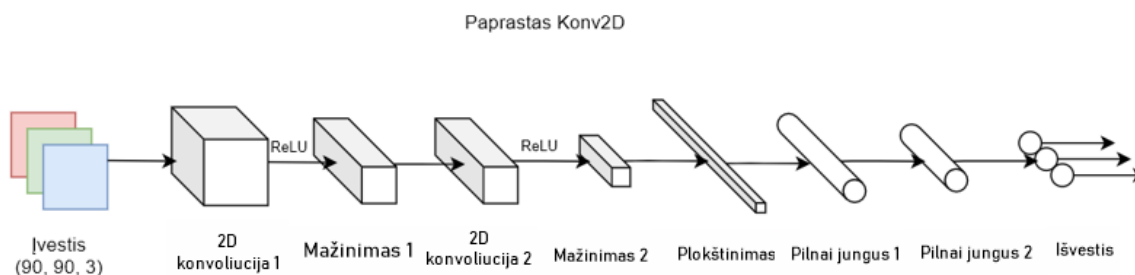
Neuroninis tinklas, dirbantis su statine vaizdine medžiaga, bando klasifikuoti duomenis pagal vieną statinį vaizdo įrašo kadrą. Tai pats paprasčiausias būdas klasifikuoti vaizdinius duomenis, tačiau gali būti nepakankamas jeigu klasifikavimui reikalinga vaizdo pokyčių informacija.

Numanoma, kad klasifikuojant kompiuterinių žaidimų vaizdo medžiagą vaizdų judėjimas gali nešti svarbios informacijos. Siekiant patikrinti šį teiginį reikia palyginti dviejų panašių neuroninių tinklų klasifikavimo tikslumą, kur vienas tinklas dirba su statine informacija, o kitas su dinamine.

#### 4.1.1. Paprasti konvoliuciniai sluoksniai

Statinams spalvotiems vaizdams klasifikuoti buvo sukurtas paprastas, kelių sluoksnių konvoliucinis neuroninis tinklas klasifikatorius. Tinklo modelis pavaizduotas 13 pav., o tinklo parametrai pateikti 5 lentelėje. Kiekviena lentelės eilutė aprašo vieną tinklo sluoksnį. Pirmas stulpelis nurodo sluoksnio pavadinimą. Antras stulpelis nurodo parametrų reikšmes būdingas atitinkamiems sluoksnių tipams, pvz., konvoliucijų sluoksnyje parametrai nurodo filtro dimensijas ir filtrų skaičių, o pilnai jungiame sluoksnyje – neuronų skaičių. Trečias stulpelis nurodo išvesties dydžius. Kiekvieno sluoksnio įvesties dydžiai atitinka prieš tai buvusio sluoksnio išvestį, išskyrus įvesties sluoksniuose, kur išvesties ir įvesties dydis toks pats. Toks žymėjimas naudojamas pateikti parametrus ir toliau nagrinėjamuose tinkluose.

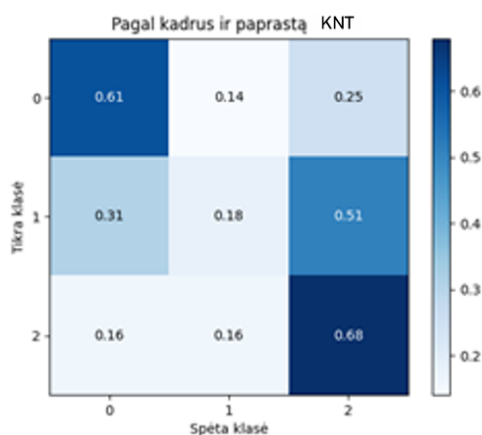
Tinklas buvo apmokytas su visais mokymosi duomenų poaibio duomenimis. Tada tinklas buvo ištestuotas su likusiais testavimo poaibio duomenimis.



13 pav. Paprasta kelių sluoksnių konvoliucinių tinklų klasifikatoriaus architektūra.

Paprastam kelių sluoksnių KNT nustatytas 49 % tikslumas klasifikuojant 3 klases. Tiksliau rezultatai pateikti 14 pav. sumaišymų matricos (angl. *confusion matrix*) forma.

Sumaišymų matricoje pavaizduota, kaip tinklo spėjimai pasiskirsto tarp tikrų ir spėjamų klasių. Eilutėse pateiktos tikrosios žaidimų klasės, o stulpeliuose tinklo spėtos klasės. Jeigu tikslumas būtų 100 %, tai visur, kur eilutėse ir stulpeliuose sutampa klasės indeksas turėtų būti įrašytas vienetas, o visi kiti langeliai turėtų įrašytus nulius. Šiuo atveju pirmoje eilutėje, pažymėtoje indeksu 0, matome vaikams skirtų žaidimų spėjimų pasiskirstymą: 61 % žaidimų buvo klasifikuoti teisingai, 14 % žaidimų klasifikuoti klaidingai, kaip paaugliams skirti žaidimai ir likę 25 % žaidimų klasifikuoti klaidingai, kaip suaugusiems skirti žaidimai. Didžiausias, 68 %, klasifikavimo tikslumas yra suaugusiųjų amžiaus klasėje pažymėtoje numeriu 2, antra pagal tikslumą yra vaikų klasė – 61 %. Taip pat matome, kad paauglių amžiaus klasė, pažymėta numeriu 1, nuspėjama mažiausiu tikslumu – tik 18 %. Daugumą (51 %) paauglių amžiui skirtų žaidimų tinklas klaidingai klasifikavo kaip skirtus suaugusiems.



14 pav. Statinius vaizdus nagrinėjančio paprasto KNT normuota sumaišymų matrica.

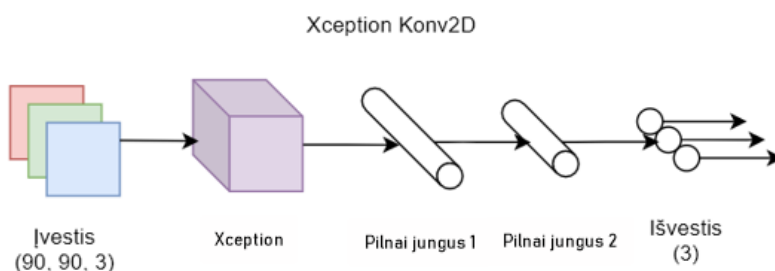
5 lentelė. Paprasto kelių sluoksnių KNT klasifikatoriaus parametrai

Pavadinimas	Parametrai	Išvestis
Ivestis	-	90, 90, 3
2D konvoliucija 1	2, 2, 128	89, 89, 128
Mažinimas 1	2, 2	44, 44, 128
2D konvoliucija 2	2, 2, 168	43, 43, 168
Mažinimas 2	2, 2	21, 21, 168
Plokštinimas	-	74088
Pilnai jungus 1	512	512
Pilnai jungus 2	256	256
Išvestis	3	3

Taigi, paprasto kelių sluoksnių KNT tikslumas, nagrinėjant statinius vaizdus, ganėtinai mažas. Nors suaugusiems skirtus žaidimus nuo vaikams skirtų žaidimų tinklas atpažino visai neblogai, visiškai nesisekė klasifikuoti paaugliams skirtų žaidimų. Taip galėjo atsitikti dėl kelių priežasčių. Viena iš jų gali būti, kad tinklas per daug paprastas, kad aptiktų paaugliams ir kitoms dvejoms amžiaus grupėms skirtų žaidimų požymių skirtumus. Pvz., smurto požymis yra būdingas tiek suaugusiems tiek paaugliams skirtiems žaidimams, tačiau paaugliams skirtuose žaidimuose smurtas yra švelnesnis, fantastiniame kontekste arba nerealistiškas, tuo tarpu suaugusiems skirtuose žaidimuose smurtas yra realistiškas arba šlykštus. Jeigu tinklas negeba atskirti smurto tikroviškumo, tai bet kokį smurtą gali būti linkęs klasifikuoti, kaip būdingą suaugusiems skirtai kategorijai. Kita galimybė yra, kad įvesties duomenyse trūko požymių, pagal kuriuos būtų galima atskirti šias klases. T. y. buvo nagrinėjami pavieniai statiniai vaizdai, todėl tikimybė, kad jie atspindės visam žaidimui būdingus požymius, pagal kuriuos buvo vertintas žaidimas, yra maža. Dar kita galimybė yra, kad paaugliams ir suaugusiems skirti žaidimai yra fundamentaliai labai panašūs. Tai reikštų, kad tikėtis aiškių požymių, pagal kuriuos būtų galima tiksliai atskirti šias dvi klases, yra beprasmiška.

#### 4.1.2. Iš anksto apmokyti konvoliuciniai sluoksniai

Patikrinkime, ar dėl klaidingo klasifikavimo kaltas per paprastas tinklas. Vietoje to, kad mokytumėme KNT išrinkti požymius pagal mūsų duomenų aibę, galime taikyti iš anksto apmokytus KNT modulius, kurie apmokyti identifikuoti objektus didelėse paveikslėlių aibėse. Tinklo architektūra pateikta 15 pav.

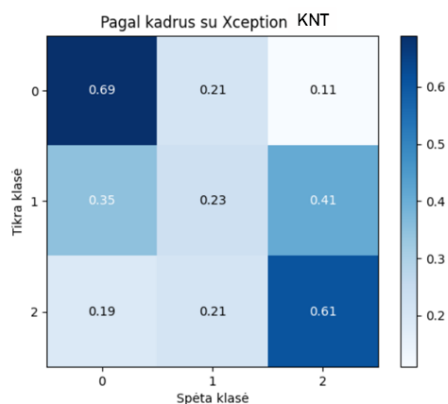


15 pav. Kelių sluoksnių konvoliucinių tinklų klasifikatoriaus architektūra su Xception moduliui.

Kaip iš anksto apmokytą tinklą naudokime Xception [Cho17]. Tinklas buvo apmokytas naudojantis *ImageNet* duomenų aibę, kur pirmieji tinklo sluoksniai išmoko išskirti požymius iš daugiau nei milijono paveikslėlių klasifikuojant juos į 1000 skirtingų klasių. Šis tinklas pasižymi paprasta ir nelabai didele architektūra, tačiau aplenkia klasifikavimo uždaviniuose tokius tinklus kaip *Inception V3*.

Naudosime tik požymius išskiriančią *Xception* tinklo dalį. Kadangi mūsų klasifikavimo uždavinys kitoks, nei *Xception* buvo apmokytas klasifikuoti, tai *Xception* jungieji sluoksniai mums netinka. Todėl iš paskutinio *Xception* požymius išrenkančio sluoksnio išvesime į mūsų jungius sluoksnius ir mokysime tinklą iš naujo. *Xception* mokytas dirbti su 299 taškų ilgio ir pločio vaizdo failais, tačiau autorių teigimu požymius išrenkantys sluoksniai turėtų veikti ir su mažesniais vaizdais. Kiti tinklo požymiai pateikti 6 lentelėje.

Tinklas su *Xception* požymių išskyrimo sluoksniais pasiekė 51 % vidutinį tikslumą. Tai 2 % didesnis tikslumas nei naudojant paprastus konvoliucijos sluoksnius. Tačiau, nors paauglių ir vaikų amžiaus klases tinklas pradėjo atskirti šiek tiek geriau, vis tiek 16 pav. matomas didelis sumaišymų skaičius. Tikėtina, kad dėl prieš tai minėtų priežasčių, statiniuose vaizduose tiesiog nėra pakankamai informacijos atskirti tokias panašias klases, todėl reikia išbandyti nagrinėti dinامينius vaizdus.



16 pav. *Xception* KNT normuota sumaišymų matrica.

6 lentelė. *Xception* KNT klasifikatoriaus parametrai

Pavadinimas	Parametrai	Išvestis
Įvestis	-	90, 90, 3
<i>Xception</i>	-	2048
Pilnai jungus 1	512	512
Pilnai jungus 2	256	256
Išvestis	3	3

## 4.2. Darbas su dinamine vaizdine medžiaga

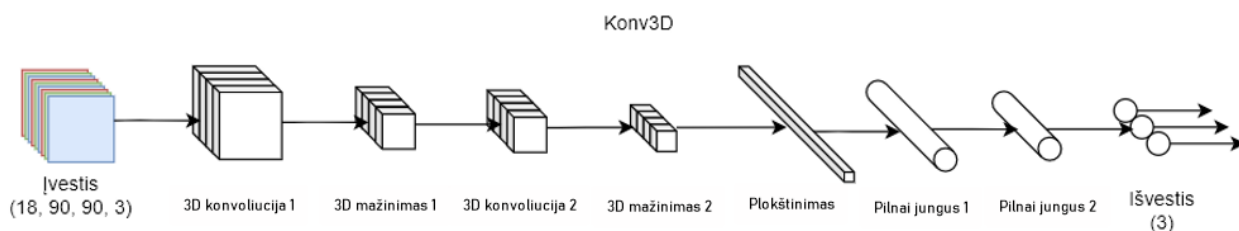
Dinaminiuose vaizduose saugoma daugiau informacijos. Juose užfiksuoti judesiai gali padėti identifikuoti įvykius ar jų požymius, kurie padės sėkmingai klasifikuoti vaizdinę medžiagą. Tačiau daugiau informacijos reiškia ir daugiau duomenų, kuriuose tinklas turi išmokti surasti reikalingą informaciją.

Šiame skyriuje nagrinėsime kelis skirtingus būdus ir nustatysime, kuris iš jų tinka geriausiai mūsų klasifikavimo uždaviniui spręsti.



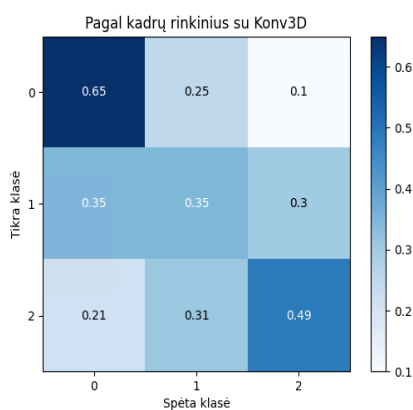
### 4.2.1. Trijų dimensijų konvoliuciniai filtrai

Vienas paprasčiausių būdu dirbti su dinaminiais vaizdais yra naudoti trijų dimensijų konvoliucinius filtrus. Laikome, kad dinaminis vaizdas tėra statinių vaizdų rinkinys, kuris surenkamas papildomoje dimensijoje. Tuomet, perkėlę esamus konvoliucinius filtrus į aukštesnę dimensiją, galime išskirti požymius iš kadrų rinkinių. Kaip pavaizduota 17 pav., toks neuroninis tinklas yra labai panašus į prieš tai naudotą statinį. Trimatis KNT mechanizmas geriausiai veikia aptinkant trumpus dinaminius požymius, kadangi filtrai apdoroja tik kelis šalia esančius kadrus.



17 pav. Trijų dimensijų KNT architektūra.

Su šiuo tinklu pasiektas 49,5 % tikslumas. Tai yra pusė procento daugiau nei statinio tinklo su paprastais KNT sluoksniais, tačiau mažiau nei su statiniais vaizdais ir *Xception* moduliu. Tačiau, 18 pav. matomas lygesnis pasiskirstymas tarp klasifikavimo tikslumų lyginant su 14 pav. ir 16 pav. Šis tinklas paaugliams skirtus žaidimus klasifikuoja 35 % tikslumu, o prieš tai buvę tinklai klasifikavo tik 18 % ir 23 % tikslumu. Šio tinklo parametrai pateikti 7 lentelėje.



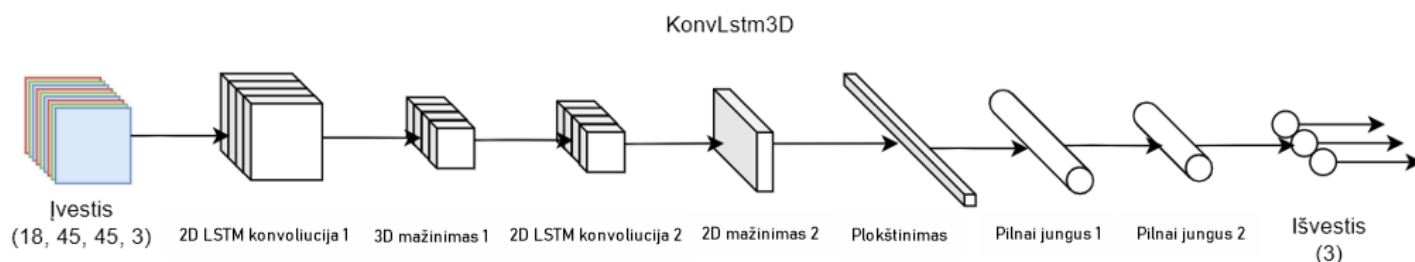
18 pav. Trimačio KNT normuota sumaišymų matrica.

7 lentelė. Trijų dimensijų KNT parametrai.

Pavadinimas	Parametrai	Išvestis
Išvestis	-	18, 90, 90, 3
3D konvoliucija 1	2, 2, 2, 64	17, 89, 89, 64
3D mažinimas 1	2, 2, 2	8, 44, 44, 64
3D konvoliucija 2	2, 2, 2, 64	6, 43, 43, 64
3D mažinimas 2	1, 2, 2	6, 21, 21, 64
Plokštinimas	-	169 344
Pilnai jungus 1	128	128
Pilnai jungus 2	64	64
Išvestis	3	3

## 4.2.2. Konvoliuciniai filtrai su ilga trumpalaike atmintimi

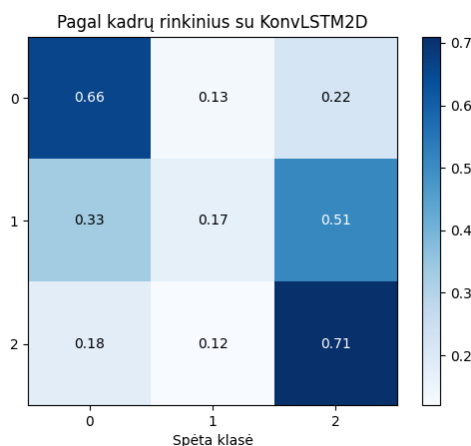
Kitas būdas apdoroti vaizdo duomenų kitimą laike yra naudojant konvoliucinius tinklus su ilga trumpalaike atmintimi (angl. *long-short term memory*, *LSTM*). Šie tinklai priima kadrų rinkinius ir saugo išrinktus požymius atmintyje. Tada požymius galima išvesti kaip vieną matricą arba padalintą į rinkinius. Tinklo architektūra pavaizduota 19 pav.



19 pav. Konvoliucinių tinklo su ilga trumpalaike atmintimi architektūra.

Tinklo architektūra labai panaši į trimačių konvoliucijų, tik vietoje trimačių konvoliucijų naudojamas vienas konvoliucijų blokas su kadrų rinkinių išėjimu, tada mažinimo sluoksnis ir antras konvoliucijos su atmintimi sluoksnis, kurio išvestis yra suagreguota.

Šis tinklas pasiekė 50,9 % tikslumą. Tai truputį geresnis tikslumas nei trimačiai konvoliuciniai sluoksniai. Tačiau 20 pav. matome, kad antros klasės klasifikavimo tikslumas yra tik 17 % – mažiausias iš visų iki šiol nagrinėtų tinklų. Tinklo parametrai pateikti 8 lentelėje.



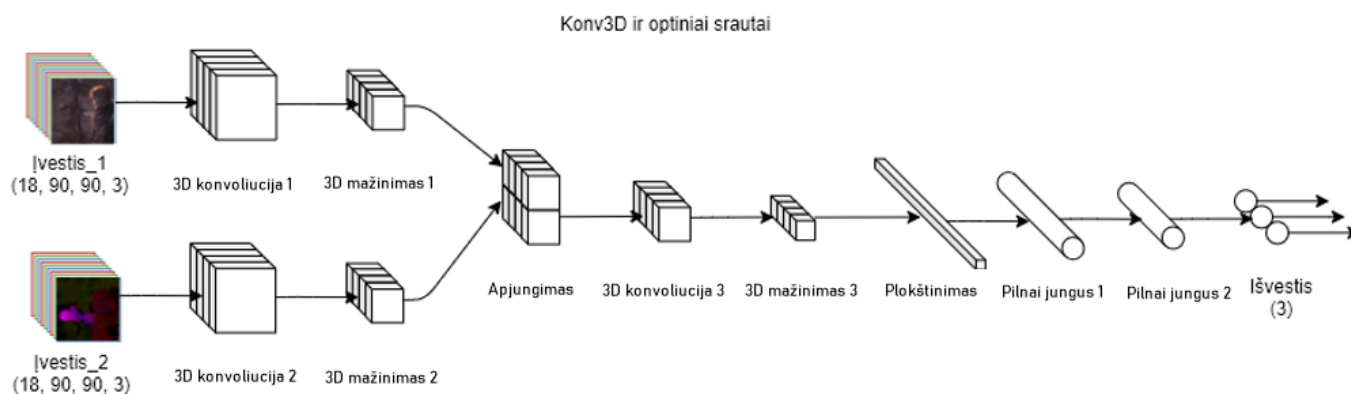
20 pav. Dvimačio KNT su ilga trumpalaike atmintimi normuota sumaišymų matrica

8 lentelė. Dvimačio KNT su ilga trumpalaike atmintimi parametrai.

Pavadinimas	Parametrai	Išvestis
Įvestis	-	18, 45, 45, 3
2D LSTM konvoliucija 1	2, 2, 2, 32	18, 43, 43, 32
3D mažinimas	2, 2, 2	9, 21, 21, 32
2D LSTM konvoliucija 2	2, 2, 2, 32	19, 19, 32
2D mažinimas	2, 2	9, 9, 32
Plokštinimas	-	2 529
Pilnai jungus 1	128	128
Pilnai jungus 2	64	64
Išvestis	3	3

### 4.2.3. Išretinti kadrai ir optiniai judesio srautai

Optiniai srautai saugo judėjimo informaciją. Kiekvienas vaizdo taškas koduoja judėjimo krypties ir greičio vektorių, apskaičiuotą pagal vaizdo įrašo kadrus. Tai teoriškai yra naudinga informacija, nes tinklui turėtų padėti išrinkti požymius apie judančius vaizdus. Tinklo architektūra pateikta 21 pav.



21 pav. Trimačio KNT su kadry ir optinių srautų įvestimi architektūra.

Optinių srautų efektyvumas priklauso nuo vaizdo įrašo medžiagos. Geriausi optinių srautų rezultatai gaunami, kai vaizdo įrašo kamera yra stabili ir juda tik objektai vaizde. Jeigu kamera nėra stabili, tai kameros judėjimas yra užfiksuojamas kaip visų vaizde esančių objektų judėjimas. Tokia informacija gali pasidaryti perteklinė ir nereikalinga. Pavyzdžiui, 22 pav. pavaizduoti dvi vaizdo įrašų ištraukos ir iš jų sugeneruoti optiniai srautai. Variante (b) matome, kad kai kamera yra stabili, tai optiniuose srautuose matomas objekto, šiuo atveju žmogaus pakeliančio ranką ir rodančio kryptį, vaizdas. Tuo tarpu variante (a) vaizdo įrašas yra iš pirmo asmens perspektyvos ir judant žaidimo veikėjui juda visa kamera, dėl to sugenerotas optinis srautas yra triukšmingas ir tiksliai nieko nesimato.



a) Nestabili kamera



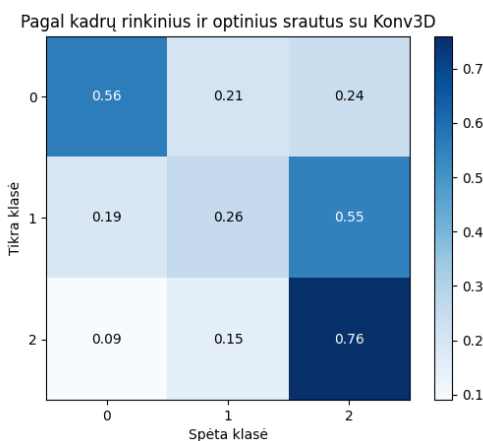
b) Stabili kamera

22 pav. Vaizdo įrašo kadrai ir atitinkami optiniai judesio laukai.

Šio tinklo tikslumas buvo 52,5 %. Tai geriausias tikslumas iš kol kas nagrinėtų tinklų. Tačiau 23 pav. sumaišymų matricioje matoma tokia pati prasto paaugliams skirtų žaidimų klasifikavimo problema, kaip ir kituose tinkluose. Teisingai suklasifikuojama tik 26 % paaugliams skirtų žaidimų, daugiau nei pusė jų (55 %) klaidingai klasifikuojami kaip skirti suaugusiems. Šio tinklo parametrai pateikti 9 lentelėje.

Taigi, nors žaidimuose vaizdo kamera dažniausiai nėra stabili ir optiniai srautai yra triukšmingi, vis tiek matomas klasifikavimo tikslumo padidėjimas lyginant su tinklais be optinių srautų. Tai rodo, kad triukšmingi optinių srautų kadrai neklaidina tinklo klasifikavimo.

9 lentelė. Trimačio KNT su optiniais srautais parametrai.



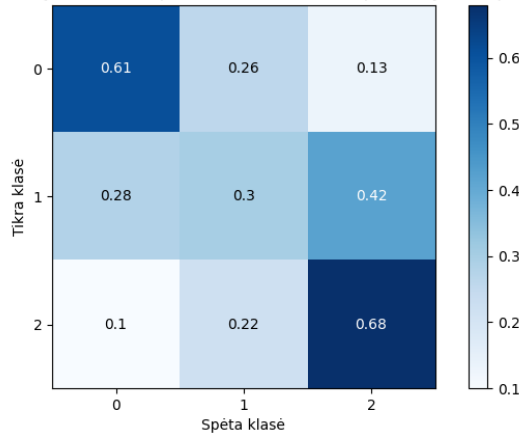
23 pav. Kadrijų su optiniais srautais ir trimačiu KNT normuota sumaišymų matrica

Pavadinimas	Parametrai	Išvestis
Įvestis 1	-	18, 90, 90, 3
Įvestis 2	-	18, 90, 90, 3
3D konvoliucija 1	2, 2, 2, 64	17, 89, 89, 64
3D mažinimas 1	2, 2, 2	8, 44, 44, 64
3D konvoliucija 2	2, 2, 2, 64	17, 89, 89, 64
3D mažinimas 2	2, 2, 2	8, 44, 44, 64
Apjungimas	-	8, 44, 44, 128
3D konvoliucija 3	2, 2, 2, 64	7, 43, 43, 64
3D mažinimas 3	1, 2, 2	7, 21, 21, 64
Plokštinimas	-	197 568
Pilnai jungus 1	128	128
Pilnai jungus 2	64	64
Išvestis	3	3

Pakeitus trimačius KNT sluoksnius į dvimatį su trumpa ilgalaikė atmintimi gautas žymiai mažesnis – 39 % tikslumas. Tai galbūt būtų galima paaiškinti tuo, kad optiniai srautai savaime saugo judėjimo laike informaciją, todėl analizuojant optinių srautų kitimą laike gauname nebe judesį, bet judesio kitimą laike, o tai gali būti perteklinė ir ne tokia naudinga informacija renkant amžiaus kategorijų požymius.

Pakeitus požymius išrenkantį sluoksnį į laike išskirstytą *Xception* gautas kol kas didžiausias – 53 % tikslumas. Tai truputį geresnis rezultatas nei tiesiog trimatis KNT. Taip pat, 24 pav. sumaišymų matricoje matome, kad gaunamas truputį didesnis tikslumas klasifikuojant antrą klasę – 30 % lyginant su trimačio KNT 26 %. Tačiau tinklas vis tiek maišo antros ir trečios klasės įrašus (42 % paaugliams skirtų žaidimų klaidingai klasifikuojami kaip skirti suaugusiems), kas buvo būdinga ir visiems kitiems iki šiol nagrinėtiems tinklams.

Pagal kadru rinkinius ir optinius srautus su laike paskirstytu Xception



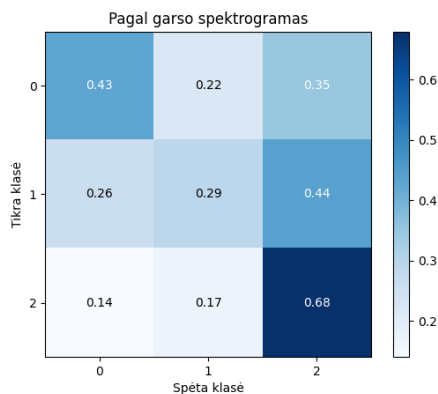
24 pav. Kadru ir optinių srautų su laike paskirstytu Xception normuota sumaišymų matrica.

### 4.3. Darbas su garsine medžiaga

Klasifikuoti garsą galima jį pavertus į spektrogramas. Kadangi spektrogramos fiksuoja garso signalų dažnių intensyvumus laike, tai nereikia sudėtingo modelio, kuris išrinktų požymius iš duomenų kintančių laike. Uždavinys tampa labai panašus į statinio vaizdo kadro klasifikavimą.

#### 4.3.1. Spektrogramos ir konvoliuciniai filtrai

Panaudojus tinklą, kuris buvo naudotas statinių paveikslukų gavimui, su keliomis modifikacijomis gautas 46,8 % klasifikavimo tikslumas. Pagal 25 pav. sumaišymų matricą matome, kad tinklas linkęs žaidimus klasifikuoti kaip skirtus suaugusiems – 30 % vaikams skirtų žaidimų ir 46 % paaugliams skirtų žaidimų neteisingai klasifikuoti, kaip skirti suaugusiems. Tai reiškia, kad tinklas turi didelį jautrumą suaugusiems skirtų žaidimų garsų požymiams aptikti. Tinklo parametrai pateikti 10 lentelėje.



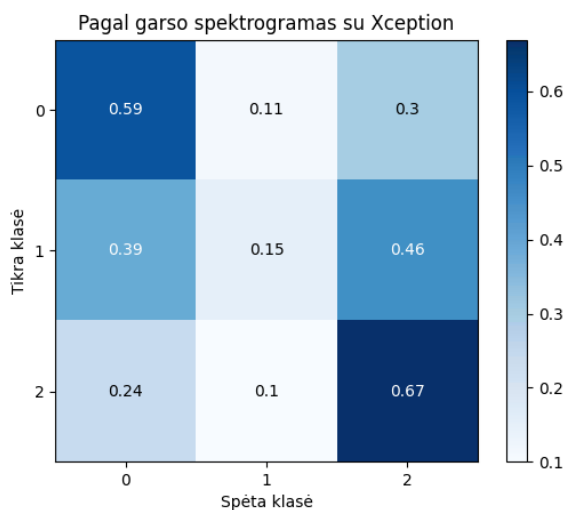
25 pav. Spektrogramos su KNT normuota sumaišymų matrica.

10 lentelė. Spektrogramos su KNT tinklo parametrai.

Pavadinimas	Parametrai	Išvestis
Įvestis	-	64, 64, 3
2D konvoliucija 1	2, 2, 64	64, 64, 128
2D konvoliucija 2	2, 2, 64	64, 64, 128
2D mažinimas 1	2, 2	32, 32, 128
2D konvoliucija 3	2, 2, 64	32, 32, 128
2D mažinimas 2	2, 2	16, 16, 128
2D konvoliucija 4	2, 2, 64	14, 14, 128
2D mažinimas 3	2, 2	7, 7, 128
Plokštinimas	-	6 272
Pilnai jungus 1	256	256
Pilnai jungus 2	256	256
Išvestis	3	3

### 4.3.2. Iš anksto apmokyti neuroniniai tinklai

Pakeitus požymius išskiriančius KNT sluoksnius į iš anksto apmokytą *Xception* gautas šiek tiek didesnis – 47 % tikslumas. Kaip ir su statiniais vaizdo kadrais, matoma, kad *Xception* išskiria pakankamai bendrus požymius, kad galėtų nagrinėti ir ne vaizdinių objektų duomenis. Tačiau, kaip matome 26 pav. sumaišymų matricoje, kaip ir daugelis kitų tinklų, šis taip pat sunkiai identifikuoja antrąją klasę – paaugliams skirti žaidimai teisingai klasifikuojami tik 15 % visų žaidimų.



26 pav. Spektrogramos su *Xception* normuota sumaišymų matrica

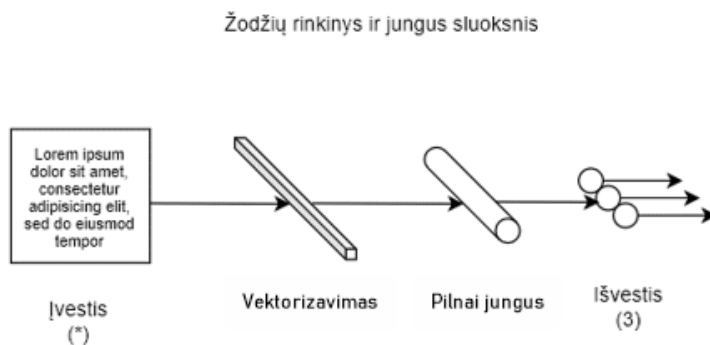
### 4.4. Darbas su tekstine medžiaga

Tekstinių aprašų tipas yra labiausiai nutolęs nuo kitų duomenų tipų sprendžiant šį uždavinį. Vaizdo ir garso įrašai teikia realią informaciją iš žaidimo, pagal kurią buvo nustatytas amžiaus cenzas. Tuo tarpu teksto aprašymas yra autorių interpretacija ir žaidimo atpasakojimas. Iš vieno pusės tai gali reikšti, kad teksto aprašymas tiksliai neatitiks to, koks aprašomas žaidimas yra iš tikrųjų. Iš kitos pusės, jeigu aprašymas tikslus, jis gali labai koncentruotai nusakyti koks yra žaidimo turinys, ko trumpas žaidimo vaizdo ir garso ištraukos gali ir neatspindėti.

Tekstinę medžiagą nesunku komponuoti su kitais duomenų tipais. Požymiai iš teksto išrenkami nepriklausomai nuo kitų duomenų, todėl užtenka išrinktus požymius apjungti vėlesniuose, pilnai jungiuose klasifikavimo sluoksniuose. Todėl, nepriklausomai nuo teksto individualaus klasifikavimo tikslumo, verta pabandyti tekstą prijungti prie bendro tinklo.

#### 4.4.1. Žodžių rinkinys

Paprasčiausias būdas dirbti su tekstiniais duomenimis yra paverčiant tekstą į žodžių rinkinio vektorių (angl. *bag-of-words*). Iš visų unikalių tekstuose esančių žodžių yra sudaromas žodynas. Kiekvienas galimas aprašo žodis yra laikomas atskiru įvesties požymiu. Tada žodžių rinkinio vektoriaus ilgis sutampa su žodyno ilgiu ir kiekvienas vektoriaus elementas žymi atitinkamo žodžio pasikartojimo dažnį. Tinklo architektūra pateikta 27 pav.

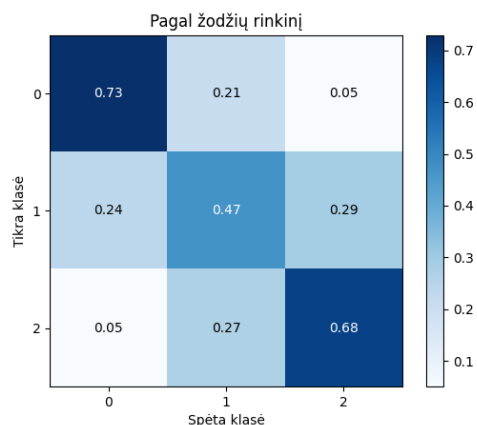


27 pav. Žodžių rinkinio tinklo architektūra.

Žaidimų duomenų aibėje tekstinius aprašus sudaro 10 352 unikalių žodžių. Taigi, tinklo įvestis yra vienmatė matrica su 10 352 elementų, kur kiekvienas elementas nurodo, kiek kartų pasikartojo atitinkamas žodis. Taip pat mokymosi metu buvo išimti žodžiai, kurie pasikartoja žaidimo pavadinime. Taip siekiama užtikrinti, kad tinklas nesimokytų žaidimo pavadinimų, bet ieškotų kitų amžiaus cenzą apibūdinančių žodžių.

Šis tinklas pasiekė gana didelį 62,7 % tikslumą. Tinklo parametrai pateikti 11 lentelėje. Pagal 28 pav. pavaizduotą sumaišymų matricą taip pat matome, kad paaugliams skirtų žaidimų klasifikavimo tikslumas yra geriausias (47 %) iš visų iki šiol nagrinėtų tinklų.





11 lentelė. Žodžių rinkinio tinklo parametrai.

Pavadinimas	Parametrai	Išvestis
Įvestis	-	*
Vektorizavimas	-	10 352
Pilnai jungus	128	128
Išvestis	3	3

28 pav. Tinklo su žodžių rinkinio vektoriumi normuota sumaišymų matrica

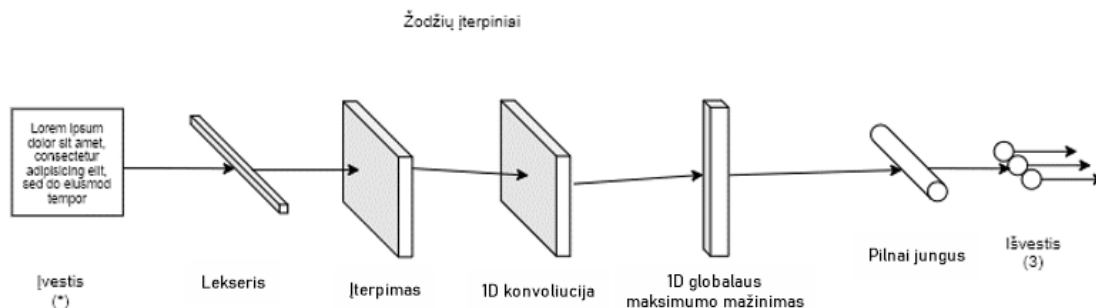
Bendrai didesnę tikslumą pasiekti lyginant su vaizdą ar garsą nagrinėjančiais tinklais tikriausiai padeda teksto aprašų reprezentatyvumas žaidimo turiniui, kurio trumpose žaidimo vaizdo ir garso įrašuose yra mažiau. Pvz., siaubo žaidimo *Outlast* tekstiniame apraše yra tokios raktines frazės kaip „apleistame beprotnamyje laukia siaubas“ ir „bauginanti linija tarp mokslo ir religijos“, kurios stipriai indikuoja, kad žaidimas yra skirtas suaugusiems. Tuo tarpu analizuojamuose vaizdo įrašuose matosi tik apleistas pastatas (29 pav.), o garso įrašuose girdisi sunkus kvėpavimas ir siaubo žaidimams būdinga muzika, kas yra daug daugiau duomenų (lyginant vaizdo ir garso įrašą su tekstu), bet reikalingos informacijos tankis yra mažesnis.



29 pav. Siaubo žaidimo *Outlast* vaizdo įrašo kadras.

## 4.4.2. Žodžių įterpiniai

Žodžių įterpinių tinkle įvestis tampa  $n$ -mačių vektorių rinkiniu. Kiekvienas vektorius atitinka vieną žodį. Tai reiškia, kad tinklo įvesties ilgis apribojamas žodžių skaičiumi. Jeigu įvesties tekstas trumpesnis, tai matrica užpildoma nuliais. Tinklo architektūra pateikta 30 pav.

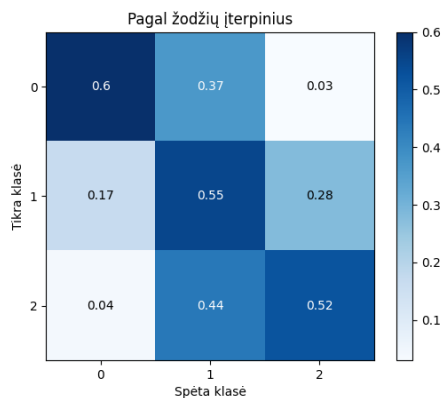


30 pav. Žodžių įterpinių tinklo architektūra.

Prieš įvedant tekstą, kaip ir žodžių rinkinių tinkle, yra sudaromas žodžių žodynas. Tačiau toliau žodžių indeksas žodyne yra susiejamas su  $n$ -mačiu vektoriumi, kuris atitinka tą žodį visų žodžių erdvėje. Šioje erdvėje susiję žodžiai yra linkę būti arti vienas kito. Taip siekiama efektyviai užkoduoti tekste esančią informaciją.

Žodžiai paverčiami į  $n$ -matį vektorių naudojantis įterpinių matricą. Matricos reikšmės arba svoriai gali būti apskaičiuojama mokymosi metu, arba galima naudoti iš anksto apmokytą įterpinių matricą, kuri efektyviai paskirsto žodžius erdvėje. Įvesties tekstą konvertavus į vektorių matricą galima iš šios matricos ištraukti požymius pasinaudojus vienmačiais konvoliuciniais sluoksniais.

Igyvendinus šia architektūrą tinklas parodė 55,6 % tikslumą. Taip pat, pagal 31 pav. sumaišymų matricą matome, kad tinklas efektyviau, t. y. 55 % tikslumu, nei kitų duomenų tipų tinklai nustato paaugliams skirtų žaidimų klases. Tinklo parametrai pateikti 12 lentelėje.

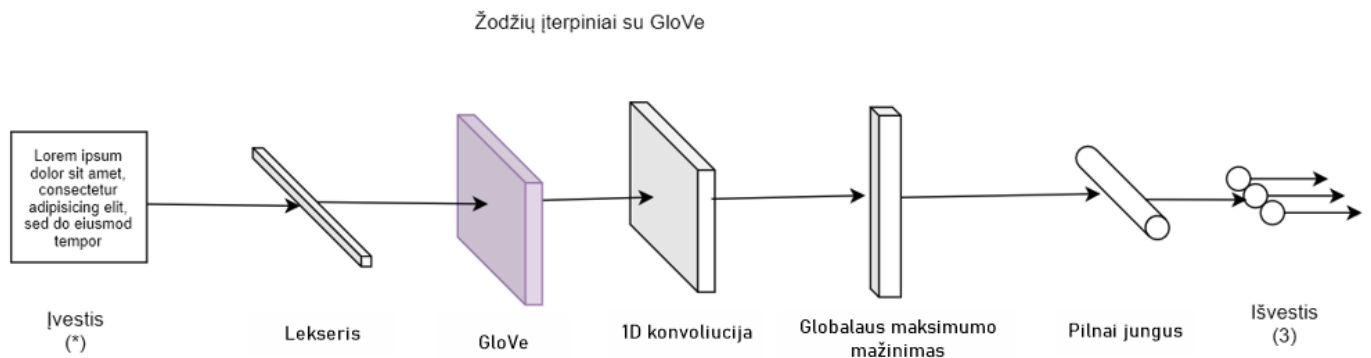


12 lentelė. Žodžių įterpinių tinklo parametrai.

Pavadinimas	Parametrai	Išvestis
Įvestis	-	100, 50
1D konvoliucija 1	3, 128	98, 128
1D globalaus maksimumo mažinimas 1	2, 2	128
Pilnai jungus 1	128	128
Pilnai jungus 2	128	128
Išvestis	3	3

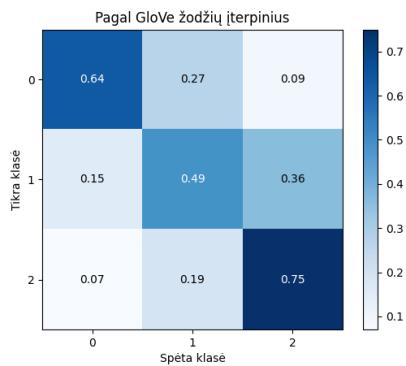
31 pav. Žodžių įterpinių tinklo normuota sumaišymų matrica.

Pakeitus tinklo įterpinių sluoksnį iš anksto apmokytu sluoksniu *GloVe* [PSM14], gauname 32 pav. pavaizduotą architektūrą. Sluoksnis *GloVe* yra apmokytas pagal didelį kiekį žodžių, todėl įvesties vektoriai turėtų perteikti tikslesnę informaciją, nei mokymosi metu sukurta įterpinių matrica.



32 pav. Tekstą apdorojančio tinklo su *GloVe* įterpimo sluoksniu architektūra

Tinklas, su *GloVe* žodžių įterpiniais, pasiekė 62,7 % tikslumą. Tinklo sumaišymų matrica pateikta 33 pav. Tai geresnis tikslumas nei įterpimo sluoksnį mokant nuo pradžių, tačiau toks pats kaip naudojant žodžių rinkinio metodą. Bet *GloVe* įterpinių tinklas yra dvigubai mažesnis – jis turi 581 tūkst. parametru, o rinkinio – 1,3 mln. Tinklo parametrai pateikti 13 lentelėje.



33 pav. Žodžių įterpinių GloVe tinklo normuota sumaišymų matrica

13 lentelė. Tekstą apdorojančio tinklo su GloVe įterpimo sluoksniu parametrai

Pavadinimas	Parametrai	Išvestis
Įvestis	-	100
GloVe	-	100, 50
1D konvoliucija	3, 128	98, 128
1D globalaus maksimumo mažinimas	2, 2	128
Pilnai jungus	128	128
Išvestis	3	3

## 4.5. Apibendrinimas

Dirbant su vieno tipo duomenimis didžiausią tikslumą žaidimų amžiaus cenzo klasifikavimui duoda tekstiniai aprašai. Didžiausias tikslumas pasiektas su žodžių rinkinio ir *GloVe* žodžių įterpimo tinklais – 62,7 %. Tikėtina, kad tekstinius aprašus klasifikuojantis tinklas pasiekė didesnę tikslumą nei vaizdo ar garso įrašus nagrinėjantys tinklai dėl tekstiniuose aprašuose esančios reprezentatyvesnės informacijos viso žaidimo požymiams išrinkti, tuo tarpu nagrinėjami multimedijos duomenys yra trumpos ištraukos ir ne taip gerai reprezentuoja viso žaidimo požymius.

Nagrinėjant vaizdo įrašus buvo nustatyta, kad didžiausią tikslumą klasifikuojant amžius cenzą teikia laike išskirstytas *Xception* tinklas su vaizdo įrašo kadrais ir optiniais srautais. Pasiektas tikslumas yra 53 %. Remiantis šiuo tinklo tikslumu jis bus panaudojamas vaizdo medžiagai analizuoti komponuotame tinkle.

Nagrinėjant garsus buvo nustatyta, kad iš anksto apmokytas *Xception* požymių išrinkimo modelis taip pat rodo geresnius rezultatus nei paprasti KNT. Buvo pasiektas 47 % tikslumas ir tinklas bus naudojamas komponuotame modelyje.

Taip pat, buvo matomas sudėtingumas atskirti paaugliams skirtų žaidimų amžiaus grupę. Tai ypač matėsi su multimedijos duomenimis dirbančiuose tinkluose. Tokio rezultato galima tikėtis, atsižvelgiant į tai, kad yra natūralu, kad vaikams skirtuose žaidimuose yra daugiau skirtumų lyginant juos su suaugusiems skirtais žaidimais, nei paaugliams skirtuose žaidimuose. Tai ir trumpų multimedijos duomenų nerepresentatyvumas tikriausiai ir lėmė prastą šios klasės klasifikavimo tikslumą.

## 5. Neuroninių tinklų komponavimas

Siekiant padidinti klasifikavimo tikslumą galima bandyti komponuoti kelis tinklus. Tokiu atveju vieno objekto klasifikavimo metu skirtingi tinklai gali priimti įvairesnius duomenis (pvz., vaizdą, garsą ir tekstinius aprašus).

Vienas būdas apjungti kelis tinklus klasifikuojant vieną objektą yra apjungti skirtingų tinklų gautus rezultatus. Rezultatus galima apjungti naudojant pasirinktą apjungimo funkciją. Tokiu atveju atsiranda reikalavimas, kad visų apjungiamų tinklų išeityse naudojamos vienodos klasės. Mūsų uždavinio atveju toks apribojimas problemų nesukelia, nes visus tris duomenų tipus klasifikuojantys tinklai klasifikuoja į tas pačias tris klases.

Kitas būdas apjungti tinklus yra naudojantis apjungimo sluoksniais. Tokiu atveju iš kelių tinklų yra sudaromas vienas tinklas. Visi tinklai išlaiko atskiras įvestis, į kurias paduodami atitinkamų duomenų tipų duomenys, ir atskirus požymius išrenkančius sluoksnius. Toliau išrinkti požymiai yra apjungiami apjungimo sluoksniais ir klasifikuojami bendruose klasifikavimo sluoksniuose, kurių pabaigoje gaunamas bendras rezultatas.

Abu apjungimo sluoksniai gali turėti ir pranašumų ir trūkumų. Apjungiant tik skirtingų tinklų rezultatus mažai padidinamas sprendimo sudėtingumas – naudojami tie patys tinklai, pridedama tik apjungimo funkcija. Apjungiant kelis tinklus į vieną sudėtingėja tinklo architektūra – reikia apjungti išrinktus požymius ir juos pateikti į bendrą klasifikavimo tinklą. Taip pat apjungti požymiai turi didesnę duomenų apimtį, o daugiau duomenų reiškia sudėtingesnį tinklo mokymąsi. Iš kitos pusės, apjungus visus požymius, klasifikavimo sluoksniai turi pilnesnį klasifikuojamo objekto vaizdą ir tai gali padidinti klasifikavimo tikslumą, lyginant su rezultatų apjungimu, kur kiekvienas rezultatas yra nustatytas pagal atskirus požymius.

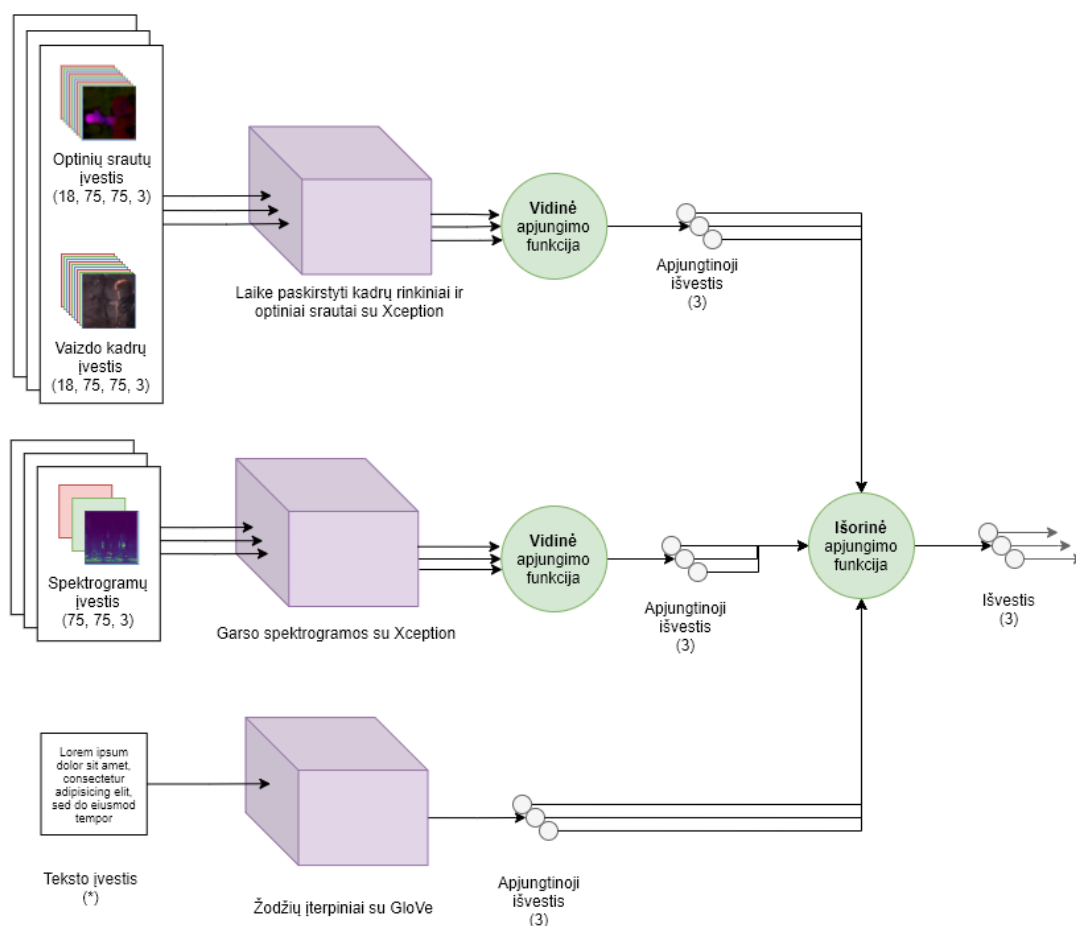
Tolimesniuose skyreliuose pateikti aptartų apjungimo būdų įgyvendinimai, jų rezultatai ir palyginimai. Naudoti tinklų komponentai remiasi prieš tai skyriuje atrinktais geriausiai tinklais kiekvienam duomenų tipui klasifikuoti.

### 5.1. Kelių tinklų rezultatų apjungimas funkcija

Apjungiant skirtingų tinklų rezultatus reikia pasirinkti apjungimo funkciją. Ši funkcija priims kelių skirtingų tinklų rezultatus ir išves vieną rezultatą, kuris bus galutinis klasifikavimo spėjimas. Taip pat, vaizdo ir garso ištraukų tinklai dirba su keliomis to pačio žaidimo ištraukomis, todėl

reikalinga antra funkcija, kuri apjungia kiekvienos ištraukos spėjimą į vieną rezultatą. Ištraukų rezultatų apjungimo funkciją vadinsime vidine, o tinklų rezultatų apjungimo – išorine.

Bus naudojami trys geriausių tikslumą turintys tinklai kiekvienam duomenų tipui su vidinėmis ir išorinėmis apjungimo funkcijomis. Tinklų diagrama pavaizduota 34 pav. Vaizdo medžiagai naudojamas tinklas priimančias vaizdo kadrus ir optinius srautus su laike išskirstytu *Xception* požymių išskyrimu. Šis tinklas vykdo klasifikavimą pagal 10 žaidimo ištraukų ir visų ištraukų įvertinimai apjungiami vidinėje apjungimo funkcijoje (diagramoje pažymėta žaliai). Garsui naudojamas tinklas priimančias garso spektrogramas su *Xception* požymius išrenkančiais sluoksniais. Su šiuo tinklu taip pat naudojama 10 garso ištraukų ir jų rezultatai apjungiami vidinėje apjungimo funkcijoje. Tekstui naudojamas tinklas su žodžių įterpiniais ir iš anksto paruošta įterpinių matrica *GloVe*. Trys gauti rezultatai apjungiami išorinės apjungimo funkcijos ir gaunamas galutinis visų tinklų klasifikavimo spėjimas. Ši konfigūracija išbandoma su skirtingomis vidurkio ir maksimumo funkcijų variacijomis.



34 pav. Kelių tinklų rezultatų apjungimas funkcijomis.

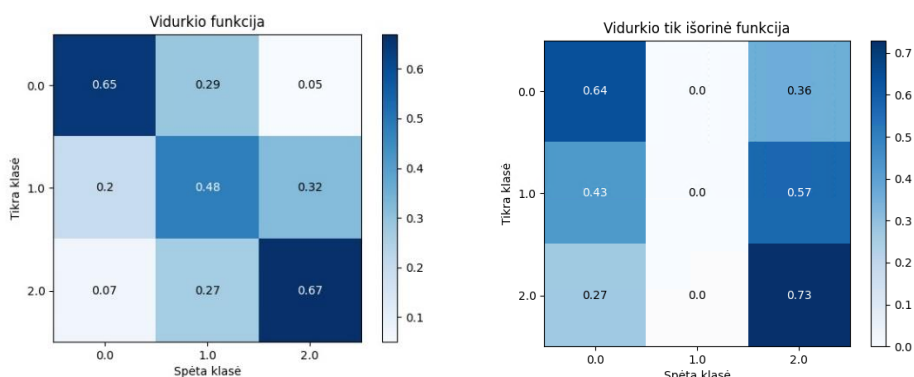
Vidinė apjungimo funkcija reikalinga tinklų įtakai suvienodinti. Vaizdo ir garso tinklai pagal 10 skirtingų žaidimo ištraukų sukuria 10 rezultatų. Jei nebūtų vidinės funkcijos, tai išorinė funkcija apjungtų rezultatus iš 21 rezultato: 10 vaizdo, 10 garso ir 1 teksto. Tada skaičiuojant tokias funkcijas, kaip, pavyzdžiui, vidurkį, tinklai turėtų ne vienodą įtaką – santykinai skaičiuojant vaizdo ir garso tinklai turėtų 10 kartų didesnę įtaką galutiniam rezultatui, nei tekstinius duomenis nagrinėjantis tinklas. Vidinė funkcija iš pradžių apjungia visų ištraukų rezultatus į vieną, t. y. iš 10 rezultatų padaro 1 rezultatą, ir tada išorinei funkcijai lieka tik trys rezultatai – iš kiekvieno tinklo po vieną. Vidurkio funkcijos atveju toks apjungimas suteikia vienodą svorį kiekvienam tinklui. Toliau pateikiamas vidurkio funkcijos pritaikymas su vidine funkcija ir be jos, ir palyginami rezultatai.

Pirmiausiai nagrinėjama vidurkio funkcija. Vidinės apjungimo funkcijos atveju vidurkio funkcija skaičiuoja kiekvienos klasės tikimybės vidurkį iš dešimties ištraukų spėjimų. Išorinės apjungimo funkcijos atveju skaičiuojamas kiekvienos klasės tikimybės vidurkis iš trijų rezultatų, gautų iš kiekvieno tinklo nepriklausomai. Taip pat atliktas atskiras skaičiavimas be vidinės funkcijos, kai išorinė apjungimo funkcija skaičiuoja vidurkį tarp 21 gauto rezultato.

Vidurkio apjungimo funkcija su vidine ir išorine funkcija pasiekė 60 % klasifikavimo tikslumą. Išorinė vidurkio funkcija be vidinės funkcijos pasiekė 45,7 % klasifikavimo tikslumą. Tikslūs apjungimo būdų rezultatai pateikti sumaišymų matricos pavidalu 35 pav., dalyje (a) yra rezultatai su vidine ir išorine vidurkio funkcija, o dalyje (b) – tik su išorine vidurkio funkcija. Apjungimas su vidine ir išorine funkcija pasiekė žymiai didesnę tikslumą, nei apjungimas tik su išorine funkcija. Nagrinėjant sumaišymų matricą matome, kad apjungimas tik su išorine funkcija visiškai neklasifikavo antrosios klasės, t. y. paaugliams skirtų žaidimų. Tai žymiai sumažino šio apjungimo tikslumą. Lyginant kitų klasių klasifikavimo tikslumus taip pat matome, kad apjungimo tik su išorine funkcija tikslumas vis tiek yra mažesnis pirmos klasės atveju (65 % prieš 64 %) ir tik šiek tiek didesnis trečios klasės atveju (67 % prieš 73 %).

Apjungiant rezultatus gaunamas prastesnis tikslumas, kai žaidimo ištraukos turi vienodą svorį lyginant su kitais tinklais. Tokių rezultatų galima tikėtis atsižvelgiant į tai, kad ne visos žaidimo ištraukos yra reprezentatyvios visam žaidimui ir dėl to turi mažą vertinimo tikslumą. Todėl priskiriant vienodą klasifikavimo rezultato svorį nereprezentatyviai vaizdo ar garso ištraukai su mažu klasifikavimo tikslumu ir visam žaidimo aprašui, kuris turi didesnę tikslumą, gauname prastesnius rezultatus.

Pridėjus vidinė apjungimo funkciją gaunami geresni rezultatai. Vidinė funkcija apskaičiuoja ištraukų vidurkius kiekviename tinkle ir tik tada apskaičiuoti vidurkiai yra dar kartą apjungiami su kitų tinklų rezultatais. Taip skaičiuojant rezultatų vidurkį teksto aprašas įgyja tiek pat svorio, kiek visos vaizdo arba garso ištraukas kartu.

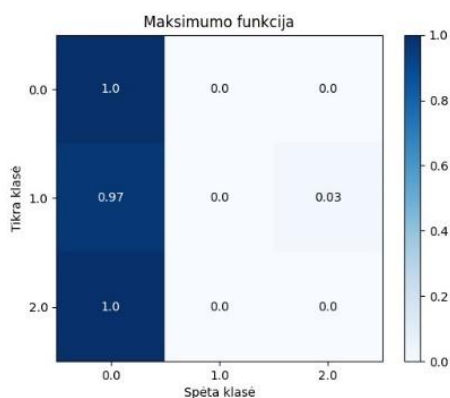


a) vidinė ir išorinė vidurkio funkcija

b) išorinė vidurkio funkcija, be vidinės

35 pav. Tinklų su vidurkio apjungimo funkcija normuotos sumaišymų matricos

Vidurkio funkciją pakeitus į maksimumo funkciją gaunami prasti rezultatai. Vidinės funkcijos atveju maksimumo funkcija išrenka didžiausią tikimybę kiekvienai klasei iš visų ištraukų. Išorinės funkcijos atveju taip pačiai išrenka didžiausią tikimybę kiekvienai klasei iš visų tinklų. Pagal 36 pav. matome, kad šis apjungimo būdas beveik visus žaidimus klasifikavo kaip skirtus vaikams. Taip yra todėl, kad pritaikius šią funkciją bet kuri viena ištrauka su labai didele arba maksimalia tikimybe nustelbia bet kurį kitą klasifikavimo rezultatą, net jeigu ji yra pavienė ir išsiskiria iš daugelio kitų.

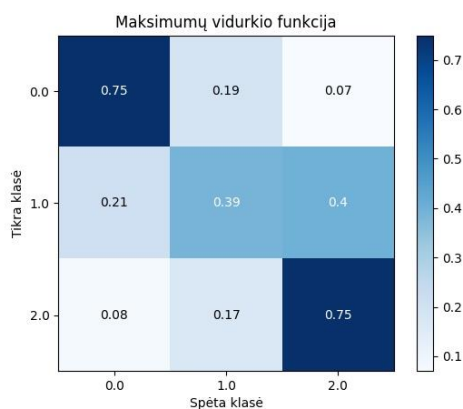


36 pav. Tinklų su maksimumo apjungimo funkcija normuota sumaišymų matrica.

Išorinę apjungimo funkciją pakeitus į vidurkio, o vidinę apjungimo funkciją palikus maksimumo gaunami geresni rezultatai. Tokiame apjungime kiekvienam tinklui iš skirtingų žaidimo



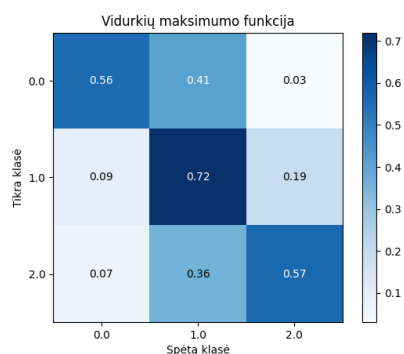
ištraukų išrenkami rezultatai su didžiausiomis tikimybės ir tada apskaičiuojamas gautų rezultatų vidurkis. Kadangi skaičiuojamas vidurkis, tai vienas atsitiktinis rezultatas su didele tikimybe negali visiškai nuspręsti galutinio rezultato. Šio apjungimo rezultatai pateikti 37 pav. sumaišymų matricioje. Bendras klasifikavimo tikslumas yra 63 %. Tai yra kol kas didžiausias pasiektas tikslumas. Lyginant su iki šiol aptartais tinklais, šis apjungimas turi ganėtinai gerą tikslumą klasifikuojant vaikams ir suaugusiems skirtus žaidimus – 75 %. Paaugliams skirtų žaidimų klasifikavimas yra mažesnis, kaip ir kituose tinkluose, tačiau nėra visiškai sumažėjęs – 39 %.



37 pav. Tinklų su maksimumo vidine funkcija ir vidurkio išorine funkcija normuota sumaišymų matrica

Sukeitus vidinę ir išorinę apjungimo funkcijas, t. y. vidinė – vidurkio funkcija, o išorinė – maksimumo funkcija, gaunamas taip pat neblogas rezultatas. Bendras apjungimo klasifikavimo tikslumas yra 62,7 %. Šiuo atveju apskaičiuojamas vidurkis tarp ištraukų klasifikavimo rezultatų ir išrenkamas rezultatas su didžiausia tikimybe. Toks skaičiavimas taip pat neleidžia pavienėms ištraukoms sugadinti vertinimo, kadangi jos įtraukiamos į bendrą vidurkį su visomis ištraukomis. Šio apjungimo rezultatai pateikiami 38 pav. sumaišymų matricioje. Matome, kad šis apjungimas turi lygintinai didelį tikslumą klasifikuojant paaugliams skirtus žaidimus – 72 %. Tačiau klasifikatorius turi didelį jautrumą paaugliams skirtiems žaidimams, t. y. net ir tie žaidimai, kurie nėra skirti paaugliams dažnai klasifikuojami kaip skirti paaugliams. Net 41 % visų vaikams skirtų žaidimų klasifikuoti kaip skirti paaugliams ir 36 % visų suaugusiems skirtų žaidimų klasifikuojami kaip skirti paaugliams. Tai vėlgi rodo, kad paaugliams skirti žaidimai yra sunkiai atskiriami nuo kitų klasių. Todėl pradėjus daugiau žaidimų teisingai klasifikuoti kaip skirtus paaugliams, pradėdama neteisingai klasifikuoti ir kitus žaidimus kaip skirtus paaugliams, ir atvirkščiai, pradėjus geriau klasifikuoti

vaikams ir suaugusiems skirtus žaidimus, pradedama daryti klaidas klasifikuojant paaugliams skirtus žaidimus.

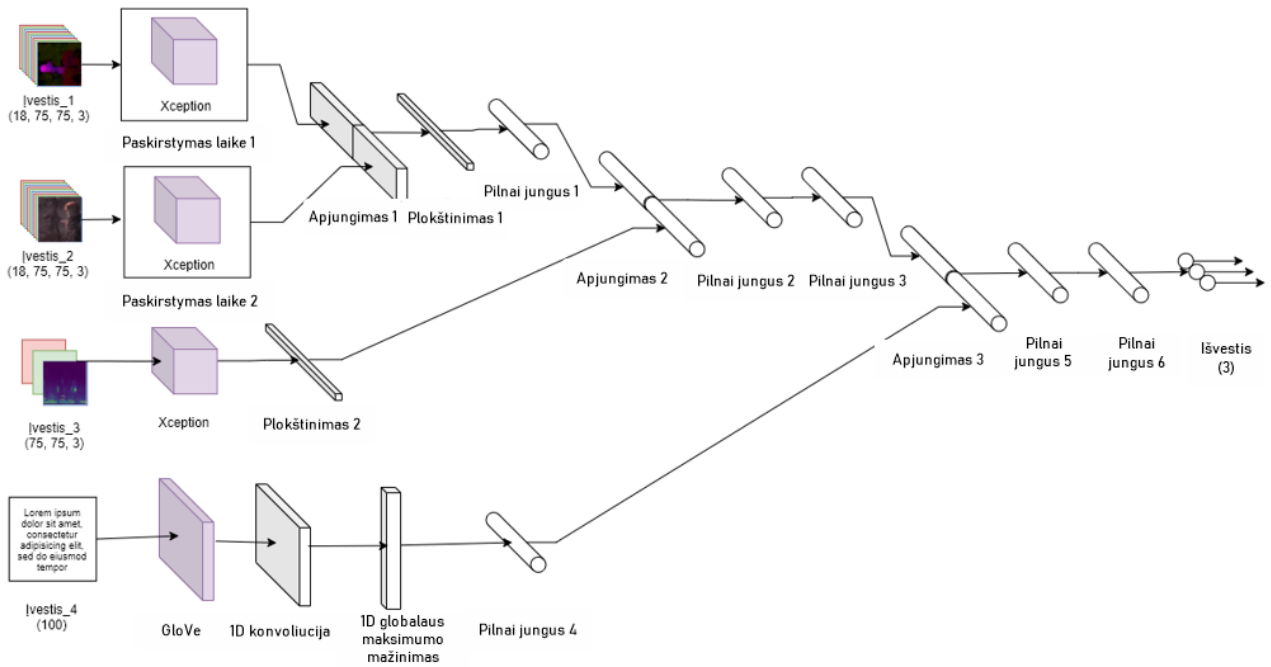


38 pav. Tinklų su vidurkių vidine funkcija ir maksimumo išorine funkcija normuota sumaišymų matrica

Tinklų rezultatų apjungimas šiek tiek padidino bendrą tikslumą palyginus su geriausiu pavienių duomenų tipo tinklu klasifikuojančiu tekstą. Geriausias bendras klasifikavimo tikslumas gautas skirtingas žaidimų ištraukas apjungiant maksimumo funkciją, o gautus skirtingų tinklų rezultatus apjungiant vidurkių funkcija.

## 5.2. Daugiarūšė komponuota tinklo architektūra

Daugiarūšis komponuotas neuroninis tinklas priima skirtingo tipo ir formos duomenis ir jais pasinaudojęs priima vieną bendrą sprendimą. Sukurto tinklo architektūra pateikta 39 pav. Šiame tinkle yra 4 įvestys: optinių srautų rinkiniai, vaizdo įrašų kadrų rinkiniai, vaizdo įrašo garso spektrogramos ir žaidimo tekstiniai aprašai.



39 pav. Daugiarūšio komponuoto tinklo architektūra

Pirmieji du įvesties sluoksniai priima optinių srautų ir atitinkamų vaizdo įrašų kadro rinkinius. Kadangi pastebėjome, kad šiam rinkiniui geriausius rezultatus rodė laike paskirstytas *Xception* požymius išskiriantis modulis, tai jį naudosime ir čia. Požymiai išskiriami kiekvienam kadru atskirai, tada sujungiami apjungimo sluoksnyje, suplokštinami ir surenkami pirmame jungiame sluoksnyje.

Trečiasis įvesties sluoksnis priima vieną statinį vaizdą, kuris vaizduoja vaizdo įrašo, iš kurio buvo paimti kadrai pirmame ir antrame įvesties sluoksniuose, garso spektrogramą. Iš spektrogramos išrinkti požymius taip pat naudojame *Xception* modulį, tik šį kartą nereikalingas paskirstymas laike, kadangi laiko duomenys spektrogramoje jau užkoduoti.

Surinkti požymiai iš optinių srautų, vaizdo įrašų ir garso spektrogramų yra apjungiami sujungimo sluoksnyje. Šie požymiai toliau apdorojami dviejuose pilnai jungiuose sluoksniuose.

Ketvirtoji įvestis priima tekstinius žaidimo aprašus. Kadangi teksto apdorojimo rezultatai buvo vienodi tiek taikant teksto įterpimą su *GloVe*, tiek taikant žodžių rinkinių vektorius, tai pasirinkta naudoti teksto įterpimą su *GloVe*, kadangi jis turėjo dvigubai mažiau požymių ir todėl bus lengviau apmokomas. Įvedamas tekstas paverčiamas į žodžių indeksus pagal sudarytą žodyną ir vykdomas tekstų įterpimas pagal *GloVe* iš anksto apmokytą teksto įterpimo matricą. Teksto ilgis apribojamas iki 100 žodžių ir visi žodžiai paverčiami į 50-mačius vektorius. Gautoje žodžių matricoje vienmatis

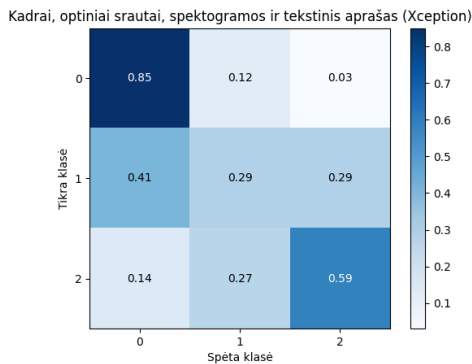
konvoliucijos sluoksnis ieško požymių. Surinkti požymiai sumažinami globalaus maksimumo sluoksnyje ir apdorojami jungiame sluoksnyje.

Tada vykdomas paskutinis požymių apjungimas, kur sujungiami jau kartu apdoroti garso ir vaizdo požymiai su teksto aprašo požymiais. Gautas požymių rinkinys dar kartą apdorojamas dviejuose pilnai jungiuose sluoksniuose ir išvedamas klasifikavimo spėjimas kategorizuotoje matricoje.

Gautame modelyje yra 68 mln. parametrų. Tačiau tik 5,6 mln. parametrų yra mokomi, kadangi naudojame iš anksto apmokytus modelius. Tai stipriai pagreitina ir taip lėta didžiulio neuroninio tinklo mokymosi procesą. Viena šio tinklo mokymo iteracija, kai tinklas pereina per visus mokymosi duomenis, truko apie dvi valandas. Architektūros parametrai pateikti 14 lentelėje.

Komponuotas neuroninis tinklas pasiekė 57,7 % tikslumą. Tai geresnis tikslumas nei pavieniai multimediją nagrinėjantys tinklai. Tačiau mažesnis tikslumas nei tik su tekstu dirbančių tinklų ar su funkcijomis apjungtų tinklų. Taip pat, pagal 40 pav. sumaišymų matricą matome, kad tinklas sunkiai klasifikuoja paaugliams skirtus žaidimus. Tinklo parametrai pateikti 14 lentelėje.

Tinklas yra didelis ir didžioji dalis tinklo išrinkinėja požymius iš žaidimo vaizdo ir garso ištrauktų. Kadangi ištraukos yra trumpos – 3 sekundžių ilgio – tai nebūtinai turi reikalingus požymius amžiaus klasei nustatyti. Dėl to, pasitaikius prastai ištraukai, didžioji dalis tinklo nebegali rasti reikalingų požymių ir galiausiai gaunamas mažesnis tikslumas nei tiesiog tekstinius duomenis apdorojančio tinklo, kuris dirba su labai reprezentatyvia informacija, pilna daug požymių.



40 pav. Komponuoto tinklo normuota sumaišymų matrica

14 lentelė. Daugiarūšio komponuoti tinklo parametrai.

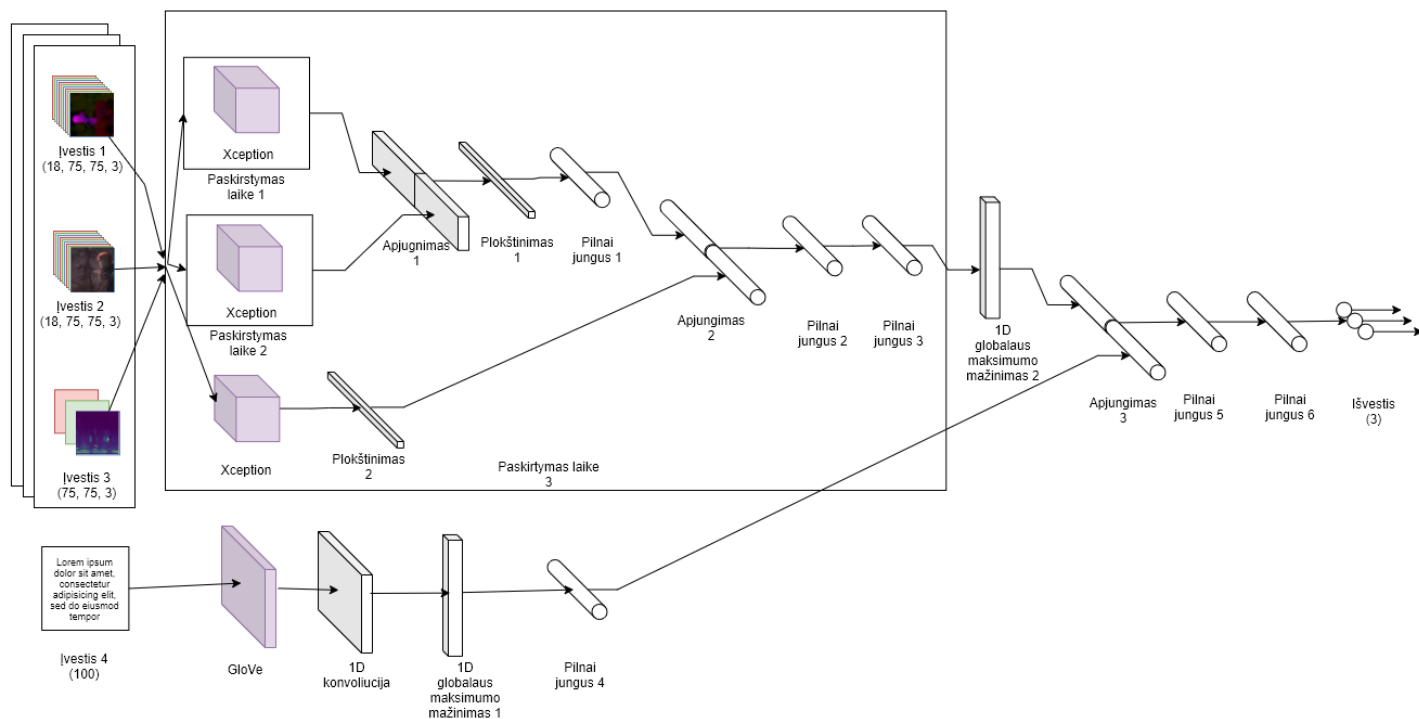
Pavadinimas	Parametrai	Išvestis
Įvestis 1	-	18, 75, 75, 3
Paskirstymas laike 1	<i>Xception</i>	18, 2048
Įvestis 2	-	18, 75, 75, 3
Paskirstymas laike 2	<i>Xception</i>	18, 2048
Apjungimas 1	Paskirstymas laike 1 Paskirstymas laike 2	18, 4096
Plokštinimas 1	73728	73728
Pilnai jungus 1	128	128
Įvestis 3	-	75, 75, 3
<i>Xception</i>	2048	2048
Plokštinimas 2	2048	2048
Apjungimas 2	2112	2112
Pilnai jungus 2	128	128
Pilnai jungus 3	128	128
Įvestis 4	-	100
<i>GloVe</i> 1	-	100, 50
1D konvoliucija 1	3, 128	98, 128
1D globalaus maksimumo mažinimas 1	128	128
Pilnai jungus 4	128	128
Apjungimas 3	256	256
Pilnai jungus 5	128	128
Pilnai jungus 6	64	64
Išvestis	3	3

### 5.3. Pakartotinė multimedijos įvestis

Viena didžiausių problemų amžiaus cenzo klasifikavime yra įvesties duomenų reprezentatyvumas. Tekstinis aprašas įprastai apima visą žaidimą, kadangi jame nupasakojame žaidimo turinį, veiksmus, kuriuos atlieka žaidėjas ir panašiai. Todėl su tekstinių aprašų duomenimis reprezentatyvumo problemos nėra. Tačiau multimedijos įvestis, šiuo metu, priima tik trumpą žaidimo ištrauką, kurioje gali nepasitaikyti jokių požymių, kurie lemia žaidimo amžiaus cenzo kategoriją.

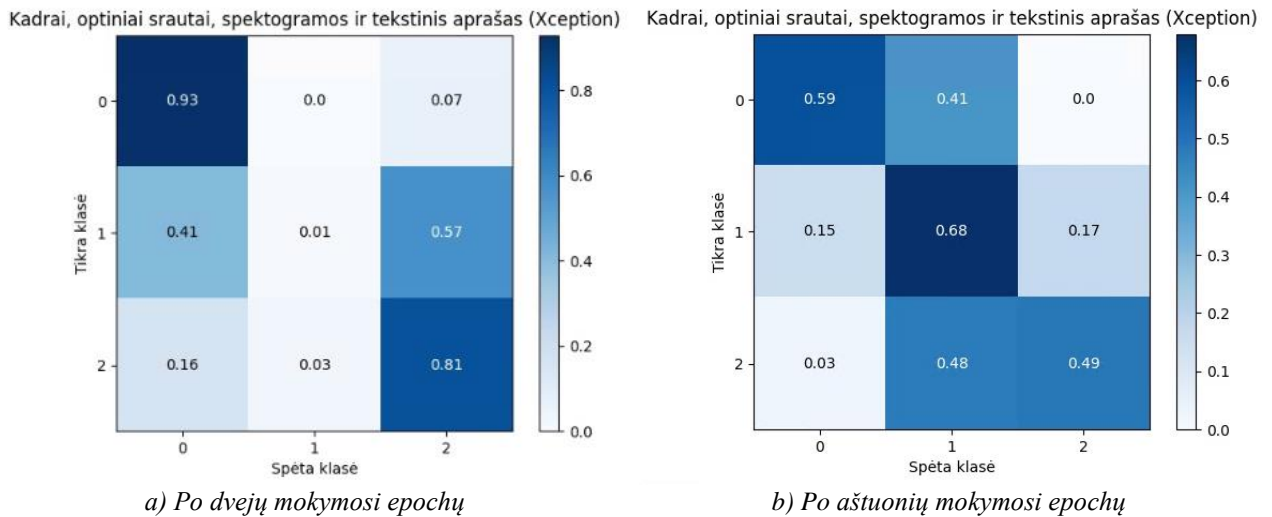
Siekiant padidinti multimedijos įvesties reprezentatyvumą pakoreguotas komponuoto tinklo multimedijos įvesties sluoksnis. Sluoksniai nuo pirmųjų trijų įvesčių (optinių srautų, vaizdo įrašų kadru ir garso signalo spektrogramos) iki apjungimo su tekstinių aprašų tinklu yra iškeliami į išskirstymo laike sluoksnį. Atlikus šį pakeitimą, multimedijos įvestis galės būti įvedama pakartotinai, imant vis kitą žaidimo ištrauką. Išskirstymo laike sluoksnis perduos kiekvieną įvestį į tinklą ir surinktus

požymius apjungs į išvesties matricą. Apdorojus visą vaizdinę medžiagą, surinkti požymiai bus praleidžiami per mažinimo sluoksni ir sujungiami kartu su teksto aprašo požymiais tolimesniam klasifikavimui atlikti, kaip ir ankstesnėje tinklo architektūroje. Atnaujinta architektūra pavaizduota 41 pav.



41 pav. Kompuoto tinklo architektūra su pakartotine įvestimi.

Atlikus šiuos pakeitimus, tinklas mokomas iš naujo jam paduodant 10 skirtingų žaidimo ištraukų, kur kiekviena ištrauka trunka 3 sekundes ir kartu paduodant žaidimo tekstinį aprašą. Dėl 10 kartų padidėjusio multimedijos įvesties duomenų kiekio tinklo mokymas stipriai sulėtėjo.



42 pav. Komponuoto tinklo su pakartotine įvestimi normuotos sumaišymų matricos

Komponuotas tinklas su pakartotine įvestimi pasiekė 58,7 % tikslumą. Tai 1 % didesnis tikslumas nei tinklo be pakartotinės įvesties, tačiau tai nėra geriausias tikslumas lyginant su kitais tinklais.

Verta pastebėti, kad mokymosi metu tinklas turėjo dvi stadijas, kurios atskleidžia, kad sunkumas atskirti paaugliams skirtus žaidimus vis dar išlieka. Pagal 42 pav. sumaišymų matricas matome, kad (a) tinklo variantas turėjo labai didelį tikslumą atskirti vaikams skirtus žaidimus nuo suaugusiems skirtų žaidimų. Vaikams skirtus žaidimus atpažino 93 % tikslumu, o suaugusiems skirtus žaidimus – 81 % tikslumu. Tokio didelio tikslumo šiose klasėse neturėjo joks kitas tinklas. Tačiau šis tinklas visiškai negalėjo atskirti paaugliams skirtų žaidimų ir dėl to vidutinis tinklo tikslumas buvo nukritęs iki 58,3 %.

Po dar šešių mokymosi epochų, kaip pavaizduota 42 pav. (b), matome, kad tinklas bandė tiksliau atspėti paaugliams skirtą klasę – tikslumas iš 1 % pakilo iki 68 %. Tačiau dėl to nukentėjo kitų klasių klasifikavimo tikslumas. Vaikams skirtų žaidimų klasifikavimo tikslumas nukrito nuo 93 % iki 59 %, o suaugusiems skirtų žaidimų klasifikavimo tikslumas nukrito nuo 81 % iki 49 %.

## Rezultatai ir išvados

Šiame darbe sprendžiamas naujas ir sudėtingas klasifikavimo uždavinys – nustatyti kompiuterinio žaidimo amžiaus cenzo kategoriją. Taip pat sukurta bendra ESRB ir PEGI vertinimų duomenų aibė, skirta mokyti dirbtinius neuroninius tinklus spręsti šį uždavinį. Išnagrinėti panašius klasifikavimo uždavinius sprendžiantys tinklai, o pagal juos įgyvendinti ir komponuojant bei modifikuojant sukurti nauji tinklai, sprendžiantys šį uždavinį. Labiausiai pasižymėjusių tinklų rezultatai pateikti 15 lentelėje, paryškintos reikšmės – didžiausios kiekviename stulpelyje.

15 lentelė. Pasižymėjusių tinklų klasifikavimo tikslumai.

Tinklas	Klasifikavimo tikslumas	Vaikams skirtų žaidimų	Paaugliams skirtų žaidimų	Suaugusiems skirtų žaidimų	Bendras
Vaizdo ir optinių srautų su laike išskirstytu <i>Xception</i>		61 %	30 %	68 %	53,0 %
Garso spektrogramų su <i>Xception</i>		59 %	15 %	67 %	47,0 %
Tekstinių aprašų su <i>GloVe</i>		64 %	49 %	75 %	62,7 %
Apjungtų tinklų pagal maksimumų vidurkius		75 %	39 %	75 %	<b>63,0 %</b>
Apjungtų tinklų pagal vidurkių maksimumus		56 %	<b>72 %</b>	57 %	62,7 %
Daugiarūšis komponuotas tinklas po 2 mokymosi epochų		<b>93 %</b>	1 %	<b>81 %</b>	58,3 %
Daugiarūšis komponuotas tinklas po 8 mokymosi epochų		59 %	68 %	49 %	58,7 %

Uždavinys pasirodė tikrai sudėtingas, norimos klasifikuoti klasės yra panašios, reikiamų apdoroti duomenų kiekis yra didelis, o reprezentatyvios informacijos tankis mažas.

### Rezultatai:

1. Sukurta 121 GB dydžio duomenų aibė tinkama neuroninių tinklų apmokymui spręsti amžiaus cenzo klasifikavimo uždavinius. Aibėje yra daugiau nei tūkstantis žaidimų su tekstiniais aprašais ir meta duomenimis, daugiau nei 11 tūkst. vaizdo įrašų ištraukų, iš kurių kiekvienam sugeneruota garso spektrograma, taip pat sugeneruota virš 200 tūkst. vaizdo įrašų kadrų su atitinkamais optiniais srautais.
2. Įgyvendinti 12 tinklų klasifikuojančių kompiuterinius žaidimus į amžiaus cenzo kategorijas pagal vaizdo, garso arba tekstinius duomenis.
3. Sukurtos ir įgyvendintos 2 tinklų architektūros, kurios yra pritaikytos klasifikuoti žaidimus į amžiaus cenzo grupes pagal skirtingus duomenų tipus:
  - a. Kelių tinklų architektūra su maksimumų vidurkio apjungimo funkcijomis pasiekia didžiausią – 63 % bendrą tikslumą.
  - b. Daugiarūšė tinklo architektūra su pakartotina įvestimi pasiekia didžiausius tikslumus specifiškai vaikų ir suaugusiųjų klasėms: 93 % tikslumą klasifikuojant žaidimus skirtus vaikams, 81 % tikslumą – suaugusiems. Tačiau,



dėl sunkumų atskiriant paaugliams skirtus žaidimus, bendras šio tinklo tikslumas yra ganėtinai mažas – 58,3 %.

4. Nustatyta įvairių duomenų tipų įtaką žaidimų amžiaus cenzo klasifikavimui.

Išvados:

1. Iš anksto su daug vaizdų ir klasių apmokytas tinklas *Xception* gali išskirti labai universalius požymius, kurie tinka analizuoti garso informaciją, optinių judesio srautų informaciją ir vaizdo informaciją sprendžiant sudėtingus klasifikavimo uždavinius.
2. Tekstiniai aprašai yra labai reprezentatyvūs sprendžiant amžiaus cenzo klasifikavimo uždavinį. Juose pateikiami apibendrinti viso žaidimo turinio aprašai, todėl lyginant su vaizdo ir garso duomenimis, ne didelis tekstinių duomenų kiekis ir paprastas tinklas pasiekia geresnį klasifikavimo tikslumą, nei bet kuris kitas duomenų tipas.
3. Dėl didelio duomenų kiekio sumažėjusį vaizdo ir garso duomenų reprezentatyvumą sprendžiant žaidimų amžiaus cenzo klasifikavimo uždavinį galima taisyti naudojant pakartotinę įvestį. Apdorojant kelis vaizdo ir garso duomenų rinkinius ir juos apjungiant globalaus maksimumo mažinimo sluoksniu, gaunami reprezentatyvesni duomenys ir pagerinamas klasifikavimo tikslumas. Vietoje vienos ištraukos tinklui paduodant 10 ištraukų tinklo bendras tikslumas padidėjo 1 %. Paduodamų ištraukų kiekį riboja tik resursų kiekis.
4. Optinių srautų naudojimas padeda išskirti reikalingus duomenis iš vaizdinių duomenų ir taip pagerinti klasifikavimo tikslumą net ir tada, kai vaizdą užfiksavusi kamera nėra visada stabili ir optiniai srautai yra triukšmingi.
5. Komponuoti tinklai rodo geresnius rezultatus, nei pavieniai tinklai. Didžiausias bendras tikslumas pasiektas apjungus tinklų rezultatus su maksimumų vidurkio funkcija – 63 %. Didžiausi specifinių klasių tikslumai pasiekti su daugiaryšiu komponuotu tinklu, vaikų žaidimų tikslumas – 93 %, suaugusių – 81 %. Nei vienas pavienis tinklas neturėjo tokio didelio tikslumo. Tai rodo, kad sujungus skirtingų duomenų tipų požymius klasifikavimo sluoksniai gali priimti tikslesnius sprendimus.
6. Kompiuterinių žaidimų amžiaus cenzo klasifikavimas yra sudėtingas uždavinys. Ypatingai sudėtinga klasifikuoti paaugliams skirtus žaidimus, kadangi šie yra tarpinė kategorija tarp vaikams ir suaugusiems skirtų žaidimų ir dėl savo panašumų dažnai klaidingai klasifikuojami kaip priklausantys vienai iš šių kategorijų.

## Literatūra

- [AAM+19] K. M. Alomari, A. Q. Alhamad, H. O. Mbaidin, S. Salloum. *Prediction of the digital game rating systems based on the ESRB*, 2019, *Opcion*, 35(Special Issue 19), p. 1368–1393.
- [ACG+19] S. Amiriparian, N. Cummins, M. Gerczuk, S. Pugachevskiy, S. Ottl, B. Schuller. “Are You Playing a Shooter Again?!” Deep Representation Learning for Audio-based Video Game Genre Recognition. *IEEE Transactions on Games*, 2019, p. 1 – 11.
- [BBR+18] V. Batchu, V. Battu, M. K. Reddy, R. Mamidi. “How to rate a video game?” – A prediction system for video games based on multimodal information. [žiūrėta 2021-05-16]. Prieiga per internetą: <<https://arxiv.org/pdf/1805.11372.pdf>>
- [Cho17] F. Chollet. Xception: Deep learning with depthwise separable convolutions, 2017, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [CPH+17] E. Cakir, G. Parascandolo, T. Heittola, H. Huttunen, T. Virtanen. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio Speech and Language Processing vol 25*, 2017, p. 1291 – 1303.
- [DFZ+14] C. Ding, S. Fan, M. Zhu, W. Feng, B. Jia. Violence detection in video by using 3D convolutional neural networks. *Lecture Notes in Computer Science*, 2014, p. 551 – 558.
- [DZW+15] Q. Dai, R. W. Zhao, Z. Wu, X. Wang, Z. Gu, W. Wu, Y.G. Jiang. Fudan-Huawei at MediaEval 2015: Detecting Violent Scenes and Affective Impact in Movies with Deep Learning. [žiūrėta 2021-05-16]. Prieiga per internetą: <<http://www.yugangjiang.info/publication/MediaEval2015-Fudan-Huawei.pdf>>
- [HBC+18] S. P. Hernandez, V. Bulitko, S. Carleton, A. Ensslin, T. Goorimoorthee. Deep learning for classification of speech accents in video games. *CEUR Workshop Proceedings, vol 2282*, 2018. [žiūrėta 2021-05-16]. Prieiga per internetą: <<https://pdfs.semanticscholar.org/c03b/2403b20d89b2f109909f1e677c32f600a3aa.pdf>>
- [LBH15] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature* 521(7553), 2015, p. 436-444.

- [MM15] A. Maulidyani, R. Manurung. Automatic identification of age-appropriate ratings of song lyrics. *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, vol 2*, 2015, p. 583 – 587.
- [MSC+13] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 2013, p. 1389 – 1399.
- [NHV+15] J. Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici. Beyond short snippets: Deep networks for video classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, p. 4694 – 4702.
- [PSM14] J. Pennington, R. Socher, C. D. Manning. Glove: Global Vectors for Word Representation. *Conference: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, p. 1532 – 1543.
- [Sch14] J. Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks* 61, 2014, p. 86 – 94.
- [SL17] S. Sudhakaran, O. Lanz. Learning to Detect Violent Videos using Convolutional Long Short-Term Memory. 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017. [žiūrėta 2021-05-16]. Prieiga per internetą: <<https://arxiv.org/pdf/1709.06531.pdf>>
- [SSI+13] M. Sjöberg, J. Schlüter, B. Ionescu, M. Schedl. FAR at MediaEval 2013 Violent Scenes Detection: Concept-based Violent Scenes Detection in Movies. [žiūrėta 2021-05-16]. Prieiga per internetą: <<https://pdfs.semanticscholar.org/a360/f69cfab9c45e6726d62363a176e2a5fe04b7.pdf>>
- [STT17] A. A. M. Al-Saffar, H. Tao, M. A. Talab. Review of deep convolution neural network in image classification. *Proceeding International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications*, 2017, p. 26 – 31.
- [SZ14] K. Simonyan, A. Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. [žiūrėta 2021-05-16]. Prieiga per internetą: <<http://arxiv.org/abs/1406.2199>>