VILNIUS TECH
Vilnius Gediminas Technical University

Mokslas – Lietuvos ateitis / Science – Future of Lithuania
ISSN 2029-2341 / eISSN 2029-2252
2023 Volume 15, Article ID: mla.2023.18834, 1–6
https://doi.org/10.3846/mla.2023.18834

Electronics and electrical engineering
Elektronika ir elektros inžinerija

DINAMINIŲ LIETUVIŲ KALBOS GESTŲ ATPAŽINIMAS

Arnas KARMONAS, Andrius KATKEVIČIUS *

Vilniaus Gedimino technikos universitetas, Vilnius, Lietuva

Gauta 2023 m. kovo 26 d.; priimta 2023 m. balandžio 7 d.

Santrauka. Rankų gestų kalba yra žmonių, turinčių klausos negalią, pagrindinis įrankis savo mintims bei žinioms perteikti. Retas žmogus, neturintis klausos negalios, supranta gestų kalbą, todėl rankų gestų atpažinimo sistemų kūrimas ir tobulinimas yra aktualus šiuolaikinis uždavinys, leidžiantis padidinti žmonių su negalia bendravimo galimybes. Rankų gestų atpažinimas taip pat leidžia bekontakčiu būdu valdyti įvairius įrenginius. Straipsnyje nagrinėjami gestų atpažinimo metodai ir pasiūlytas algoritmas, leidžiantis atpažinti dinaminius lietuvių kalbos gestus. Tyrimui buvo sukurtas dinaminių gestų duomenų rinkinys, sudarytas iš vaizdo įrašų, kurių kiekvieno trukmė yra 3 sekundės. Iš viso buvo surinkta 1100 vaizdo įrašų. Duomenų rinkinį sudarė 10 klasių. Požymiams išskirti iš vaizdo įrašo kadrų buvo naudojamas pirminio apmokymo „Inception-v3“ konvoliucinis neuronų tinklas. Išskirti požymiai buvo naudojami LSTM tinklui mokyti. Apmokytas tinklas buvo testuotas su patikros bei testavimo duomenimis ir pasiekė 85 % tikslumą.

Reikšminiai žodžiai: dinaminių gestų atpažinimas, LSTM, CNN, neuronų tinklai.

Įvadas

Rankų gestai – tai vizuali žmogaus kalba, padedanti perteikti mintis, žodžius, komandas įvairiomis rankų ar pirštų kombinacijomis bei judesiu. Bendravimą gestais naudoja kalbos arba klausos negalią turintys žmonės. Pasaulio sveikatos organizacija (angl. *World Health Organization*) teigia, apie 466 mln. žmonių, t. y. beveik 5 % visos žmonijos populiacijos turi klausos negalią (World Health Organization, 2023). Įprastai bendrauti galintys žmonės retai moka gestų kalbą. Todėl žmonių su negalia integraciją siekiama gerinti įvairiais intelektiniais metodais. Gestų atpažinimo sąsajos, kaip žmogaus ir kompiuterio komunikacijos būdas, plačiai taikomos ne tik gestams atpažinti, bet ir įvairioms elektroninėms sistemoms ar įtaisams valdyti, pvz., interaktyvūs ekranai, kompiuteriai, automobilių multimedija ir kiti (Perimal et al., 2018).

Gestai skirstomi į statinius ir dinaminius. Statiniai gestai sietini su viena konkrečia užfiksuota rankos forma ir orientacija esamoje erdvėje (Al-Shamayleh et al., 2018). Pagrindiniai statinių gestų požymiai yra rankos forma ir orientacija erdvėje, o ne atliekamas judesys. Statiniai gestai dažniausiai taikomi skaičiams, pavienėms raidėms arba specifiniams žodžiams išreikšti. Pagrindiniai iššūkiai klasifikuojant statinius gestus yra skirtingas apšvietimas, sudėtingas fonas ir odos spalvų įvairovė. Dinaminiai gestai – fiksuojama skirtingų rankos formų bei orientacijos erdvėje seka (Vuletic et al., 2019). Dinaminių gestų požymiai yra ne tik rankos forma ir orientacija erdvėje, bet ir šių požymių kitimas laike. Dinaminiai gestai dažniausiai taikomi žodžiams išreikšti ir įrenginiams valdyti, siekiant palengvinti žmonių su negalia galimybes valdyti įvairius elektroninius įtaisus. Klasifikuojant dinaminius gestus susiduriama su tokiais pat iššūkiais kaip ir atpažįstant statinius gestus, tačiau kyla ir papildomų problemų. Dinaminių gestų atlikimo įvairovę labai lemia žmonių motorika. Nėra tiksliai apibrėžta, kada dinaminis gestas prasideda ir baigiasi. Taip pat reikalingi didesni aparatinės įrangos resursai, siekiant klasifikuoti gestus realiuoju laiku (Köpüklü et al., 2019; Molchanov et al., 2016). Minėti iššūkiai skatina plačiai taikyti skirtingus intelektinius metodus gestams klasifikuoti, individualiai prisitaikant prie kiekvienos užduoties.

Zakariya ir Jindal (2019) statinių gestų klasifikavimo sistemą išmaniajame įrenginyje įgyvendino naudodami atraminių vektorių klasifikatorių (AVK) (angl. *Support vector machine*, SVM). Visų pirma nuotraukose buvo išskiriama rankos sritis, kiekvienam nuotraukos taškui

*Autorius susirašinėti. El. paštas *andrius.katkevicius@vilniustech.lt*

priskiriant dvejetainę reikšmę. Vėliau gautas dvejetainis paveikslėlis buvo naudojamos AVK mokyti. Paprastai rankų gestams atpažinti taikomi konvoliuciniai neuronų tinklai (KNT). KNT yra labai našūs, geba vaizduose išskirti erdvinius požymius ir patikimai sąveikauti su kitais klasifikatoriais (Alom et al., 2019). Islam et al. (2018) statinių gestų požymius nuotraukose išskyrė taikydami KNT. Išskirti požymiai vėliau buvo naudojami AVK mokyti.

KNT taip pat gali būti taikomi kaip vientisa požymių išskyrimo ir klasifikavimo sistema. Patel et al. (2018), Bousbai ir Merah (2019) tyrimuose kūrė KNT architektūras, kurios leistų įgyvendinti gestų atpažinimą įterptinėse sistemose, tokiose kaip išmanieji telefonai. Kurhekar et al. (2019) gestų klasifikavimą įgyvendino su „ResNet-34“ pirminio apmokymo KNT, kurie pasižymi aukštu klasifikavimo tikslumu ir laimėjo 2015 m. „ILSVRC“ bei „MS COCO“ varžybas. Šio tyrimo metu autoriai pastebėjo, jog klasifikavimo spartai didelę įtaką turi apšvietimas. Panašų tyrimą atliko Hussain et al. (2017) su šiek tiek kitokios architektūros „Resnet-18“ KNT. Sistema buvo sukurta taip, kad atsakymas būtų pateikiamas tuomet, kai bent 20 iš 30 klasifikuojamų kadrų yra tos pačios klasės. Das et al. (2018), Agrawal et al. (2020) gestų klasifikavimą įgyvendino „Inception-v3“ pirminio apmokymo KNT, kuris buvo apmokytas su „ImageNet“ duomenų rinkiniu. Rafi et al. (2019) gestų klasifikavimo uždaviniui išspręsti naudojo pirminio apmokymo VGG-19 tinklą, kurio architektūrą patobulino pakeičiant paskutinius tinklo sluoksnius, nekeičiant viršutinių tinklo sluoksnių, kurie jau buvo apmokyti. Tyrime buvo surinktas statinių bengalų kalbos gestų rinkinys iš 12 581 gestų nuotraukos, kurios buvo suklasifikuotos į 38 klases. Pasiektas 89,6 % tikslumas. Autoriai kaip sistemos trūkumą išskyrė sudėtingą foną, kuris gali pabloginti klasifikavimo tikslumą.

Klasifikuojant dinaminius gestus remiamasi ne tik rankos forma bei orientacija erdvėje, bet ir judesio požymiais. Klasifikuojant dinaminius gestus taip pat dominuoja KNT taikymas. Siriak et al. (2019) autoriai gestams klasifikuoti iš vaizdo įrašų sukūrė KNT, turintį LSTM tinklo sluoksnius. KNT išskyrė erdvinius rankų požymius iš vaizdo įrašo kadrų, o LSTM sluoksniai leido išmokti išskirtų požymių sekas. LSTM tinklai yra vieni dažniausiai naudojamų tinklų rūšių, dirbant su vaizdų sekomis. Taip pat naudojami pirminio apmokymo KNT. Dinaminių gestų atpažinimo tyrime Bantupalli ir Xie (2018) kiekvieno vaizdo kadro požymius išskyrė su „Inception-v3“ architektūros tinklu. Vėliau su išskirtais požymiais buvo apmokytas LSTM tinklas. Kitame tyrime He (2019) rankoms aptikti naudojo „Faster R-CNN“ trijų dimensijų KNT bei LSTM tinklų seką. Tokioje tinklų sekoje rankų aptikimo tinklas leido sumažinti sudėtingo fono ir nereikšmingų ypatybių vaizde kiekį. Dinaminiams gestams klasifikuoti Liao et al. (2019) pristatė naują „B3D ResNet“ architektūros tinklą. Šis tinklas sudarytas iš trijų dimensijų liekamųjų (angl. *residual*) konvoliucinių sluoksnių ir dvikrypčių LSTM sluoksnių. Sukurta architektūra leido tinklui išskirti trumpos trukmės erdvinius-laikinius (angl. *spatiotemporal*) požymius iš vaizdo sekų ir analizuoti ilgos trukmės dinamines požymių sekas.

Lietuvos tyrėjai plačiai taiko neuronų tinklus įvairiems kompiuterinės regos uždaviniams spręsti. Tumas et al. (2020) atliko tyrimą pėstiesiems aptikti atšiauriomis oro sąlygomis. Matuzevičius ir Serackis (2021) atliko trimatės žmogaus galvos rekonstrukcijos tyrimą, pagrįstą artimo nuotolio vaizdo fotogrametrija, naudodamiesi išmaniuoju telefonu.

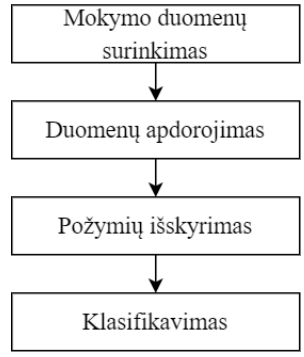
Nors yra nemažai tyrimų, klasifikuojančių statinius ir dinaminius rankų gestus pasaulyje plačiausiai naudojamomis kalbomis, tačiau yra stygius tyrimų klasifikuojant lietuvių kalbos gestus. Tokių tyrimų poreikis yra aktualus, nes kiekviena kalba turi savo individualių gestų, nes kiekviena kalba turi savo unikalių žodžių ir frazių, į kuriuos reikia atsižvelgti. Raudonis ir Jonaitis (2014) kūrė sistemą statiniams amerikiečių kalbos gestams atpažinti, kuriuos sudarė raidės ir skaičiai. Jie pasinaudojo kompiuterinės regos metodais, kad išskirtų plaštakas iš fono. Tuomet buvo naudojama diskrečioji Furjė transformacija, siekiant apskaičiuoti rankos kontūro koordinates, kuriomis buvo mokomas neuroninis tinklas. Tyrime buvo pasiektas 95,58 % klasifikavimo tikslumas.

Šiame straipsnyje pateikiama dinaminių lietuvių kalbos gestų atpažinimo metodika, kuri geba klasifikuoti 10 skirtingų gestų klasių iš vaizdo įrašų. Tyrimai atliekami su pačių autorių surinktu dinaminių lietuvių kalbos gestų duomenų rinkiniu.

1. Metodika

Šiame tyrime pateikiama vaizdais pagrįsta dinaminių lietuvių kalbos gestų atpažinimo metodika. Dinaminiai gestai, priešingai nei statiniai, atliekami judesiu. Dėl šios priežasties, klasifikuojant dinaminius lietuvių kalbos gestus, svarbu remtis ne tik erdviniais, bet ir laikiniais požymiais (1 pav.).

Dinaminių lietuvių kalbos gestų klasifikavimo metodiką sudarė 4 pagrindiniai etapai: mokymo duomenų automatizuotas surinkimas, surinktų duomenų apdorojimas, dinaminių gestų požymių išskyrimas bei dinaminių gestų klasifikavimas.



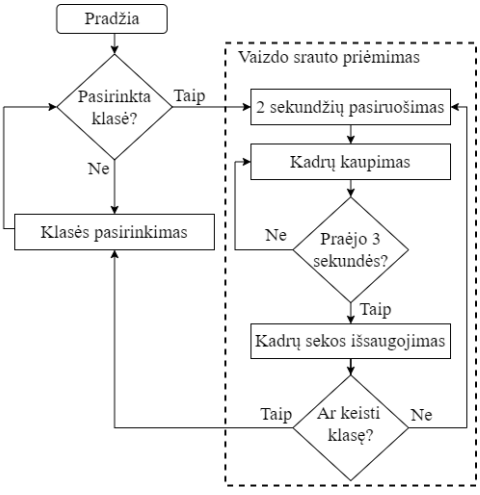
1 paveikslas. Dinaminių lietuvių kalbos gestų klasifikavimo metodikos etapai
Figure 1. Stages of dynamic Lithuanian language gestures classification methodology

1.1. Mokymo duomenų automatizuotas surinkimas

Duomenų rinkimas yra labai svarbus procesas, norint kokybiškai apmokyti dirbtinių neuronų tinklus. Tinklo apmokymo rezultatai tiesiogiai priklauso nuo duomenų rinkinio dydžio ir kokybės. Duomenų rinkimas neuronų tinklui mokyti dažniausiai yra ilga ir rankinė procedūra. Vienas pagrindinių iššūkių, renkant mokymo duomenis, dažnai būna mokymo duomenų rinkimo automatizavimas.

Automatizuotam duomenų surinkimui buvo sukurta taikomoji programa, paremta „Python“ programavimo kalba bei „OpenCV“ biblioteka. Programos veikimo algoritmas pateiktas 2 paveiksle.

Pradedant automatizuotą duomenų surinkimą visų pirma parenkama fiksuojamų rankų gestų klasė (2 pav.). Kiekvienam fiksuojamam dinaminiam rankų gestui svarbu priskirti klasę. Jei klasė yra nepasirinkta, grįžtama į klasės pasirinkimą, priešingu atveju ruošiantis atlikti gestą pradedamas rodyti kameros vaizdas. Paleidus sistemą kadrai iš rodomo vaizdo srauto pradedami kaupti tik po 2 s. Reikalingas minimalus 2 s tarpas tarp dinaminių gestų fiksavimo, siekiant pasiruošti rodyti kitą gestą. Kaupiami 640×480 rezoliucijos kadrai. Kaupimas vyksta 30 kadrų per sekundę greičiu. Praėjus 3 s, kadrų seka išsaugoma vaizdo įrašo formatu „avi“ su gesto pavadinimu bei vaizdo įrašo numeriu (pvz., Ačiū_1, Ačiū_2). Po šio etapo grįžtama į 2 s pasiruošimo etapą ir vyksta analogiškas vaizdo įrašo fiksavimas. Sukaupus norimą vaizdo įrašų skaičių, galima pakeisti rankų gestų klasę ir filmuoti kitus vaizdo įrašus. Dinaminiai rankų gestai klasifikuojami į klases, nes gestui atpažinti reikalingas prižiūrimas mokymas (angl. *Supervised learning*) ir dirbtinių neuronų tinklas turi žinoti, koks dinaminis rankų gestas pavaizduotas. Dėl šios priežasties kaupiami dinaminių lietuvių kalbos rankų gestų vaizdo įrašai buvo saugomi atskiruose aplankuose, priklausomai nuo to, kokiai dinaminių rankų gestų klasei vaizdo įrašas buvo priskirtas.



2 paveikslas. Automatizuoto dinaminių lietuvių kalbos gestų duomenų surinkimo algoritmo diagrama
Figure 2. Diagram of an automated dynamic Lithuanian language gestures data collection algorithm

Atliekant tyrimą buvo sudarytas dešimties klasių dinaminių lietuvių kalbos rankų gestų duomenų rinkinys iš 1100 vaizdo įrašų. 70 % sukauptų duomenų buvo panaudota dirbtinių neuronų tinklui mokyti. Likę 30 % dinaminių lietuvių kalbos rankų gestų buvo skirti dirbtinių neuronų tinklo patikrai. Dešimt skirtingų klasių sudarė tokie dinaminiai lietuvių kalbos rankų gestai: „sos“, „labas“, „aš“, „ačiū“, „blogai“, „gerai“, „kaip sekasi“, „viso gero“, „taip“ ir „ne“.

Vaizdo kadrų požymiams išskirti buvo naudojami konvoliuciniai neuronų tinklai, kurie yra dažniausiai naudojami dirbant su vaizdais. Išskirti požymiai buvo naudojami kaip įvestis LSTM tinklui mokyti.

1.2. Apdorojimas

Duomenų apdorojimo etape kiekvienas 3 s vaizdo įrašas išskaidomas į kadrus (nuotraukas) (1 pav.). Vėliau iš kiekvieno 3 s vaizdo įrašo kadrų sekos yra nuosekliai vienodais laiko žingsniais atrenkami 75 kadrai. Atrinkti 75 kadrai sudaro konkrečios klasės dinaminio lietuvių kalbos rankų gesto kadrų seką tinklui mokyti.

1.3. Požymių išskyrimas

Konvoliuciniai neuronų tinklai buvo pasirinkti išskirti požymius iš vaizdo įrašų kadrų sekų. Pagrindiniai šių tinklų elementai yra tokie: konvoliucinis sluoksnis, sutelkimo (angl. *pooling*) sluoksnis, aktyvavimo funkcijos bei iki galo sujungtas sluoksnis. Konvoliucinis sluoksnis yra pagrindinis šių tinklų elementas, yra atsakingas už požymių išskyrimą. Sutelkimo sluoksnis sumažina iš konvoliucinio sluoksnio gautų požymių skaičių ir leidžia naudoti mažiau parametrų mokymo metu. Aktyvavimo funkcijos lemia, kaip įvesties signalas paverčiamas į išvesties signalą ir suteikia tinklui netiesiškumo. Iki galo sujungtas sluoksnis atlieka klasifikavimą, tačiau mūsų tyrimo atveju šis sluoksnis nereikalingas, nes mums reikia tik išskirti požymius iš kadrų sekų. Klasifikavimas bus atliekamas su kitu tinklo modeliu.

Tyrime buvo pasirinktas naudoti pirminio apmokymo „Inception-v3“ konvoliucinis neuronų tinklas. Jis apmokytas su „ImageNet“ duomenų rinkiniu, turinčiu daugiau nei 1 mln. vaizdų. Pirminio apmokymo tinklai padaro uždavinio įgyvendinimo procesą greitesnį ir efektyvesnį, nes nereikia kurti tinklo nuo pat pradžių, o galima pasinaudoti esamomis standartinėmis tinklo struktūromis. Dėl šios priežasties sutaupoma skaičiavimo išteklių. „Inception“ tinklai sudaryti iš blokų. Priešingai nei tradiciniame konvoliuciniame neuronų tinkle, kur sluoksniai išdėstyti vienas po kito, „Inception-v3“ tinkle šie sluoksniai išdėstyti lygiagrečiai vienas su kitu blokuose. „Inception-v3“ tinkle iš ankstesnio bloko perduota išvestis, kitame bloke yra lygiagrečiai perduodama per kelis sluoksnius, o šių sluoksnių išvestis sujungiama ir perduodama kitam

blokui. Šiame tyrime naudojamas standartinės struktūros „Inception-v3“ tinklas.

„Inception-v3“ tinklas leido išskirti požymius iš sukauptų kiekvieno įrašo 75 kadrų sekų. Išskirti požymiai vėliau buvo naudojami LSTM tinklui mokyti.

1.4. Klasifikavimas

Požymių sekoms klasifikuoti buvo pasirinktas LSTM tinklas. Šie tinklai yra vieni dažniausių praktikoje naudojamų rekurentinių neuronų tinklų tipų. LSTM tinklai geba puikiai dirbti su duomenų sekomis ir padeda išspręsti RNN tinklų nykstančio gradiento problemą. Praktikoje šie tinklai dažnai taikomi kalbos atpažinimo, veiksmo vaizdo įraše klasifikavimo bei kitos reikšmės prognozavimo uždaviniams spręsti. Išskirtiniai šio tinklo komponentai yra įvesties, išvesties ir pamiršimo vartai, kurie kontroliuoja ilgo laiko nuoseklaus modelio klasifikavimą laiko eilučių duomenyse (Hassaballah & Awad, 2020, p. 130). Ankstesniame etape su konvoliuciniu tinklu išskirti požymiai buvo naudojami kaip įvestis LSTM tinklui mokyti.

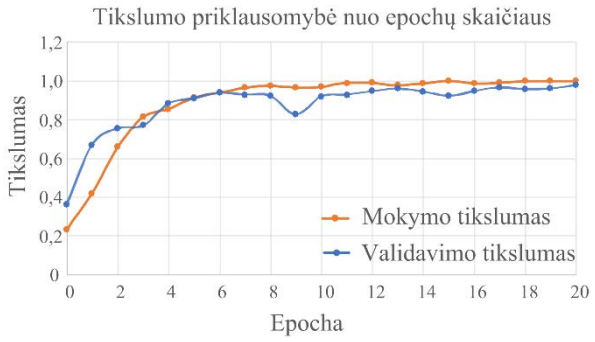
2. Rezultatai

Tyrimo metu buvo sukurtas duomenų rinkinys iš 1100 vaizdo įrašų. 770 įrašų buvo skirti mokymo tikslams. Likę 330 vaizdo įrašų buvo naudojami neuronų tinklo patikrai. Vaizdo įrašai buvo suskaidyti į kadrus. Iš kiekvieno 3 s vaizdo įrašo buvo atrinkti 75 kadrai. Požymių išskyrimui iš kiekvienos 75 kadrų sekos buvo naudotas „Inception-v3“ tinklas. Surinktos požymių sekos buvo naudotos LSTM tinklui mokyti. Mokymo metu siekiant, kad tinklas nebūtų permokytas esančiais duomenimis ir galėtų prisitaikyti prie nematytų duomenų, buvo naudojama ankstyvo sustojimo funkcija. Ši funkcija sustabdo mokymosi procesą, kai mokymo rezultatai nustoja gerėti. Apibendrinant, LSTM tinklas buvo apmokytas per 20 epochų, nes kitose 5 epochose rezultatai nebegerėjo. Mokant gauti LSTM tinklo mokymosi rezultatai atvaizduoti 3 ir 4 paveiksluose.

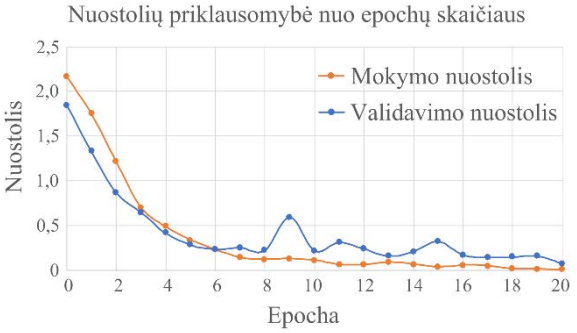
3 paveiksle galima matyti, kaip didėjo tikslumas mokymo metu. Paskutiniu mokymo žingsniu buvo pasiektas 100 % mokymo tikslumas, o patikros tikslumas siekė 97,81 %. 4 paveiksle galime matyti, kaip mažėjo nuostoliai mokymo metu. Galutiniu mokymo etapu buvo pasiekti 0,5717 % mokymo nuostoliai, o patikros nuostoliai siekė 7,4 %.

Apmokytam tinklui testuoti buvo surinktas duomenų rinkinys iš 250 vaizdo įrašų, kiekvienai klasei po 25 įrašus. Šis duomenų rinkinys kartu su patikros rinkiniu buvo pateiktas klasifikuoti. Tinklo klasifikavimo efektyvumui įvertinti su testavimo ir patikros duomenimis pateikiama klaidų matrica (5 pav.).

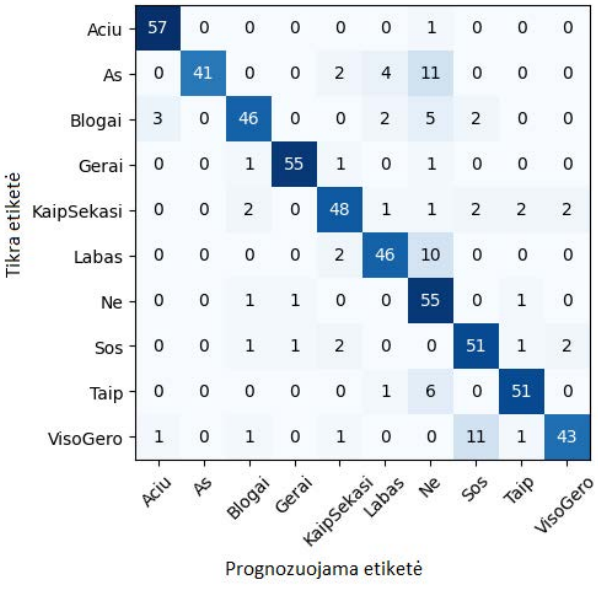
Klaidų matricoje stulpeliai atitinka tikrą gesto klasę, o eilutės nurodo prognozuojamą gesto klasę. Iš 5 paveikslo galime matyti, kad tinklas klasifikuoja gana tiksliai – 85 % tikslumu. Klasifikavimo tikslumo procentas dar išaugtų, jei būtų atmestos dvi blogiausiai klasifikuotos dinaminių rankų gestų klasės. Gestas „Aš“ 11 kartų buvo supainiotas su gestu „Ne“. O gestas „VisoGero“ buvo 11 kartų supainiotas



3 paveikslas. LSTM tinklo mokymo tikslumo priklausomybė nuo epochų skaičiaus
Figure 3. Dependence of LSTM network training accuracy on the number of epochs



4 paveikslas. LSTM tinklo mokymo nuostolių priklausomybė nuo epochų skaičiaus
Figure 4. Dependence of LSTM network training loss on the number of epochs



5 paveikslas. Dinaminių lietuvių kalbos gestų klasifikavimo klaidų matrica
Figure 5. Classification confusion matrix of dynamic Lithuanian hand gestures

su dinaminiu rankų gestu „Sos“. Reikėtų pabrėžti, kad klaidingai klasifikuoti dinaminiai lietuvių kalbos rankų gestai ne išsibarstydavo po atskiras klases padrikai, o dažniausiai būdavo sumaišomi tik su panašią judesių dinamiką turinčiais rankų gestais. Tai rodo tinklo patikimumą.

Kaip pavyzdys tiksliausiai buvo klasifikuotas dinaminis rankų gestas „Ačiū“, kuris tik kartą buvo supainiotas ir priskirtas gestų klasei „Ne“.

Išvados

Detaliai išanalizuoti įvairių kalbų dinaminių rankų gestų klasifikavimo ypatumai, taikant dirbtinių neuronų tinklus, bei atlikta jau egzistuojančių dinaminių rankų gestų duomenų rinkinių taikymo galimybių analizė.

Surinktas dinaminių lietuvių kalbos gestų duomenų rinkinys. Dinaminių lietuvių kalbos gestų požymiams išskirti panaudotas konvoliucinis neuronų tinklas. Dinaminiams lietuvių kalbos gestams klasifikuoti panaudotas LSTM tinklais. Apmokytas tinklas klasifikavo dinaminius lietuvių kalbos gestus į 10 skirtingų klasių ne mažesniu nei 85 % tikslumu.

Ateityje planuojama surinkti didesnį dinaminių lietuvių kalbos gestų duomenų rinkinį iš platesnio žmonių rato, siekiant turėti didesnę mokomųjų duomenų įvairovę. Didelį dėmesį planuojama skirti panašios judesių dinamikos rankų gestų teisingo klasifikavimo tyrimams.

Literatūra

Agrawal, M., Ainapure, R., Agrawal, S., Bhosale, S., & Desai, S. (2020, October 30–31). Models for hand gesture recognition using deep learning. In *IEEE 5th International Conference on Computing Communication and Automation* (*ICCCA*) (pp. 589–594), Greater Noida, India. https://doi.org/10.1109/ICCCA49541.2020.9250846

Alom, M. S., Hasan, M. J., & Wahid, M. F. (2019, December 24–25). Digit recognition in sign language based on convolutional neural network and support vector machine. In *2019 International Conference on Sustainable Technologies for Industry 4.0* (*STI*) (pp. 1–5), Dhaka, Bangladesh. https://doi.org/10.1109/STI47673.2019.9067999

Al-Shamayleh, A., Ahmad, R., Abushariah, M., Alam, K., & Jomhari, N. (2018). A systematic literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications*, *77*(1), 28121–28184. https://doi.org/10.1007/s11042-018-5971-z

Bantupalli, K., & Xie, Y. (2018, December 10–13). American sign language recognition using deep learning and computer vision. In *2018 IEEE International Conference on Big Data* (*Big Data*) (pp. 4896–4899), Seattle, WA, USA. https://doi.org/10.1109/BigData.2018.8622141

Bousbai, K., & Merah, M. (2019, November 24–25). A comparative study of hand gestures recognition based on MobileNetV2 and ConvNet models. In *2019 6th International Conference on Image and Signal Processing and their Applications* (*ISPA*) (pp. 1–6), Mostaganem, Algeria. https://doi.org/10.1109/ISPA48434.2019.8966918

Das, A., Gawde, S., Suratwala, K., & Kalbande, D. (2018, January 5). Sign language recognition using deep learning on custom processed static gesture images. In *2018 International Conference on Smart City and Emerging Technology* (*ICSCET*) (pp. 1–6), Mumbai, India. https://doi.org/10.1109/ICSCET.2018.8537248

Hassaballah, M., & Awad, A. I. (2020). *Deep learning in computer vision: Principles and applications* (1 ed.). CRC Press/Taylor and Francis. https://doi.org/10.1201/9781351003827

He, S. (2019, October 16–18). Research of a sign language translation system based on deep learning. In *2019 International Conference on Artificial Intelligence and Advanced Manufacturing* (*AIAM*) (pp. 392–396), Dublin, Ireland. https://doi.org/10.1109/AIAM48774.2019.00083

Hussain, S., Saxena, R., Han, X., Khan, J. A., & Shin, H. (2017, November 5–8). Hand gesture recognition using deep learning. In *2017 International SoC Design Conference* (*ISOCC*) (pp. 48–49), Seoul, Korea (South). https://doi.org/10.1109/ISOCC.2017.8368821

Islam, M. R., Mitu, U. K., Bhuiyan, R. A., & Shin, J. (2018, September 24–27). Hand gesture feature extraction using deep convolutional neural network for recognizing american sign language. In *2018 4th International Conference on Frontiers of Signal Processing* (*ICFSP*) (pp. 115–119), Poitiers, France. https://doi.org/10.1109/ICFSP.2018.8552044

Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2019, May 14–18). Real-time hand gesture detection and classification using convolutional neural networks. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition* (*FG 2019*) (pp. 1–8), Lille, France. https://doi.org/10.1109/FG.2019.8756576

Kurhekar, P., Phadtare, J., Sinha, S., & Shirsat, K. P. (2019, April 23–25). Real time sign language estimation system. In *2019 3rd International Conference on Trends in Electronics and Informatics* (*ICOEI*) (pp. 654–658), Tirunelveli, India. https://doi.org/10.1109/ICOEI.2019.8862701

Liao, Y., Xiong, P., Min, W., Min, W., & Lu, J. (2019). Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks. *IEEE Access*, *7*, 38044–38054. https://doi.org/10.1109/ACCESS.2019.2904749

Matuzevičius, D., & Serackis, A. (2021). Three-dimensional human head reconstruction using smartphone-based close-range video photogrammetry. *Applied Sciences*, *12*(1), 1–26. https://doi.org/10.3390/app12010229

Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016, June 27–30). Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) (pp. 4207–4215), Las Vegas, NV, USA. https://doi.org/10.1109/CVPR.2016.456

Patel, R., Dhakad, J., Desai, K., Gupta, T., & Correia, S. (2018, December 14–15). Hand gesture recognition system using convolutional neural networks. In *2018 4th International Conference on Computing Communication and Automation* (*ICCCA*) (pp. 1–6), Greater Noida, India. https://doi.org/10.1109/CCAA.2018.8777621

Perimal, M., Basah, S., Safar, M., & Yazid, H. (2018). Hand-gesture recognition-algorithm based on finger counting. *Journal of Telecommunication, Electronic and Computer Engineering* (*JTEC*), *10*(1–13), 19–24. https://jtec.utem.edu.my/jtec/article/view/4115

Rafi, A. M., Nawal, N., Bayev, N. S. N., Nima, L., Shahnaz, C., & Fattah, S. A. (2019, October 17–20). Image-based bengali sign language alphabet recognition for deaf and dumb community. In *2019 IEEE Global Humanitarian Technology Conference* (*GHTC*) (pp. 1–7), Seattle, WA, USA. https://doi.org/10.1109/GHTC46095.2019.9033031

Raudonis, V., & Jonaitis, D. (2014, May 8–9). Gesture sign language recognition method using artificial neural network, as a translator tool of deaf people. In *9th International Conference on Electrical and Control Technologies ECT* (pp. 13–17), Kaunas, Lithuania.

Siriak, R., Skarga-Bandurova, I., & Boltov, Y. (2019, September 18–21). Deep convolutional network with long short-term memory layers for dynamic gesture recognition. In *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications* (*IDAACS*) (pp. 158–162), Metz, France. https://doi.org/10.1109/IDAACS.2019.8924381

Tumas, P., Nowosielski, A., & Serackis, A. (2020). Pedestrian detection in severe weather conditions. *IEEE Access*, *8*, 62775–62784. https://doi.org/10.1109/access.2020.2982539

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, *129*, 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011

World Health Organization. (2023, March 25). *Deafness and hearing loss*. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Zakariya, A. M., & Jindal, R. (2019, July 6–8). Arabic sign language recognition system on smartphone. In *2019 10th International Conference on Computing, Communication and Networking Technologies* (*ICCCNT*) (pp. 1–5), Kanpur, India. https://doi.org/10.1109/ICCCNT45670.2019.8944518

RECOGNITION OF DYNAMIC LITHUANIAN LANGUAGE GESTURES

A. Karmonas, A. Katkevičius

Abstract

This paper proposes a method for automated Lithuanian hands gestures data collection and for Lithuanian hands gestures classification. The dataset of 1100 samples was collected for 10 different classes of Lithuanian hands gesture. The features of hands gestures were extracted with CNN network. The classification was made with LSTM network. The trained LSTM network classified the Lithuanian hands gestures with 85% accuracy.

Keywords: dynamic Lithuanian hands gesture recognition, LSTM, CNN, neuron networks.