

VILNIAUS UNIVERSITETAS

LAURYNAS DOVYDAITIS

LIETUVIŠKAI KALBANČIO DIKTORIAUS IDENTIFIKAVIMO
NAUDOJANT REKURENTINIUS NEURONINIUS TINKLUS
TIKSLUMO TYRIMAS

Daktaro disertacija

Technologijos mokslai, informatikos inžinerija (07 T)

Vilnius, 2018

Disertacija rengta 2013–2017 metais Vilniaus universitete.

Mokslinis vadovas

doc. dr. Vytautas Rudžionis (Vilniaus universitetas, technologijos mokslai,
informatikos inžinerija – 07 T).

PADĖKA

Nuoširdžiai dėkoju moksliniam vadovui dr. Vytautui Rudžioniui už patarimus viso studijų laikotarpio metu ir už paramą rengiant šią disertaciją.

Dėkoju disertacijos recenzentams dr. Virginijui Marcinkevičiui ir dr. Ryčiui Maskeliūnui už pastabas ir komentarus, kurių dėka šio darbo kokybė tapo geresnė.

Taip pat dėkoju dr. Kęstučiui Driauniui už paramą, paskatinimą ir palaikymą mano ilgoje mokslinėje kelionėje. Nesusitikus teisingoje vietoje ir teisingu laiku, vargu ar būčiau nuėjęs taip toli.

Už kantrybę, supratimą, bei visapusišką paramą nuoširdžiai dėkoju žmonai ir gyvenimo partnerei Lilijai. Šiame tekste neužtektų vietos apibūdinti tavo kantrybei ir supratimui.

Rengiant šią disertaciją, tiesiogiai ir netiesiogiai prie palaikymo prisidėjo nemažai žmonių. Visiems jiems noriu padėkoti už įkvėpimą ir paskatinimą nuolat judėti pirmyn.

Laurynas Dovydaitis

SANTRAUKA

Disertacijoje yra nagrinėjamas diktoriaus identifikavimo pagal balsą uždavinys. Tai yra gana sunki užduotis. Ir nors žmonės tai daro pakankamai intuityviai, tačiau prie tam tikrų sąlygų, tarkime, didelės diktorių duomenų bazės, net ir žmogui šis uždavinys gali tapti neįveikiamas. Disertacijos tyrimo objektas yra diktoriaus identifikavimas pagal akustinį balso signalą taikant rekurentinius neuroninius tinklus, o tikslas – pasiūlyti klasifikavimo metodą, skirtą lietuviškai kalbančio diktoriaus identifikavimui pagal balsą, naudojantį rekurentinius neuroninius tinklus.

Šiame darbe yra apžvelgtas asmens identifikavimo pagal balsą procesas, ir išskirti metodai, naudojami diktoriaus atpažinimui. Balso atpažinimo uždavinį sudaro keli pagrindiniai etapai. Tai diktoriaus balso apdorojimas, požymių išskyrimas, diktoriaus atpažinimo klasifikatoriaus apmokymas.

Mokslinės literatūros analizės metu nustatyta, kad dažniausiai naudojamą diktoriaus požymių, išskiriamų iš balso signalo, rinkinį sudaro melų skalės kepstro koeficientai ir jų išvestinės. Šis požymių išskyrimo metodas yra toliau naudojamas darbe, siekiant nustatyti klasifikatorių, tiksliausiai identifikuojantį lietuviškai kalbančius diktorius. Diktoriaus identifikavimo klasifikatoriaus tikslumo nustatymui yra pasiūlyta metodika, pagal kurią yra išskiriami požymiai, atrenkami tiksliausią identifikavimą suteikiantys klasifikatoriaus hiperparametrai ir atliekamas tikslumo įverčių palyginimas, bei pasiūlytas metodas lietuviškai kalbančio diktoriaus identifikavimui.

Atliktų tyrimų metu nustatyta, kad lietuviškai kalbančių diktorių identifikavimą atliekant su rekurentiniais neuroniniais tinklais ir jų dvikrypčio ilgos trumpalaikės atminties tinklo topologija, gaunamas tikslesnis diktoriaus identifikavimo rezultatas, nei naudojant paslėptuosius Markovo modelius akustiniam diktoriaus pavyzdžių modeliavimui ir jų klasifikavimui. Tyrimams buvo naudotas lietuviškai kalbančių diktorių garsynas LIEPA.

Tarpiniai darbo rezultatai buvo pristatyti 2 tarptautinėse, 5 respublikinėse konferencijose, o straipsniai publikuoti 2 recenzuojamuose periodiniuose žurnaluose, 3 recenzuojamuose konferencijų leidiniuose.

ABSTRACT

One of the main tasks of this thesis is to analyze the process of speaker identification by one's voice by classifying a set of speaker's voice features extracted from one's voice sample. This is a complex task, given the fact that even people who are usually able to do it intuitively would find it challenging if presented with an extensive speaker dataset. The objective of the present research is speaker identification from an acoustic speech signal and the aim of this work is to create a classification method in order to get higher identification accuracy for Lithuanian speakers by using recurrent neural networks.

This thesis presents methods that are used to identify a speaker by voice. The process of speaker identification consists of several stages. The first stage is the processing of a voice sample, which is then followed by a process of speaker feature extraction, and the final stage involves speaker identification classifier training.

Based on the research in the field, it was concluded that the most popular and commonly used method for speaker feature extraction is Mel frequency cepstral coefficients and its derivatives. This method is further applied in experiments conducted using a Lithuanian speaker dataset in order to determine a classifier most accurately identifying Lithuanian speaking individuals. This method includes extraction of features, selection of a classifier hyper-parameter and comparison of results. A new method for identifying Lithuanian speaking individuals is outlined.

The performed experimental research allowed concluding that for a Lithuanian speaker higher identification accuracy is achieved by using a recurrent neural network classifier with long short-term memory topology. This neural network performs better on the accuracy metric than speaker acoustic modeling with hidden Markov models. Experiments were performed on Lithuanian speaker dataset LIEPA.

Intermediary results of this thesis were presented in 2 international and 5 national conferences. Intermediary results were published in 2 peer-reviewed journals and 3 peer-reviewed conference notes.

ŽYMĖJIMAI

Simboliai	
a_{ji}	paslėpto Markovo modelio pereinamoji tikimybė, indeksai j ir i parodo perėjimo kryptį tarp diskrečiosios būsenos
$b_i(x)$	Gauso skirstiniai, kur i yra mišinio svoriai, x – D-dimensijų požymių vektorius
$b_j(o_t)$	stebėjimų tikimybinių tankio funkcija, j – paslėpto Markovo modelio būseną, t – laiko momentas, o – stebimas požymių vektorius
c_t	neuroninio tinklo paslėpto sluoksnio ilgos trumpalaikės atminties celė, kur t yra įėjimo sekos numeris
δ_t	delta koeficientas laiko momentu t
f_t	loginių pamiršimo elementų įėjimai, kur t yra įėjimo sekos numeris
$\mathcal{F}_k$	aktyvavimo funkcija
$g(x)$	tiesinio klasifikatoriaus diskriminanto funkcija
h	paslėpto sluoksnio neuronas
$\vec{h}$	paslėpto sluoksnio vektorius būsimu laiko momentu
$\tilde{h}$	paslėpto sluoksnio vektorius prieš tai buvusiu laiko momentu
i_t	loginių įėjimo elementų įėjimai, kur t yra įėjimo sekos numeris
j, k	neurono arba celės indeksas

Simboliai	
$M(f)$	dažnių skalė melais
M	super-vektorius
N_s	lango trukmė imčių skaičiumi
o_t	loginių išėjimo elementų įėjimai, kur t yra įėjimo sekos numeris
p_i	i -tasis mišinio svoris
s_k	neuroninio tinklo sklidimo taisyklė, k -tajam neuronui
t	laikas
$w(n)$	lango funkcija diskretiniam laikui n
w_{jk}	svorio koeficientas, kuris nustato j neurono poveikį k neuronui
y	išėjimo vektorius
y_j	aktyvacijos būseną kiekvienam neuronui j
Θ	delta lango dydis
μ	i -vektorius
θ_k	neuroninio tinklo įėjimo nuokrypis (angl. <i>bias</i>) k neuronui
δ	klaidos sklidimo signalas
γ	mokymosi žingsnis

Santrumpos	
AVM	atraminių vektorių mašinos (angl. <i>support vector machine</i>)
DFT	diskretinė Furjė transformacija (angl. <i>discrete Fourier transform</i>)
DRNT	dvikryptis rekurentinis neuroninis tinklas (angl. <i>bi-directional recurrent neural network</i>)
DITA	dvikryptė ilga trumpalaikė atmintis (angl. <i>bi-directional long short-term memory</i>)
DKT	diskretinė kosinusinė transformacija (angl. <i>discrete cosine transform</i>)
DP	daugiasluoksnis perceptronas (angl. <i>multi-layer perceptron</i>)
GFT	greitoji Furjė transformacija (angl. <i>fast Fourier transform</i>)
GMM	Gauso mišinių modeliai (angl. <i>Gaussian mixture models</i>)
RNT	rekurentinis neuroninis tinklas (angl. <i>recurrent neural network</i>)
ITA	ilga trumpalaikė atmintis (angl. <i>long short-term memory</i>)
JFA	jungtinių faktorių analizė (angl. <i>joint factor analysis</i>)
MSKK	melų skalės kepstriaiai koeficientai (angl. <i>mel scale cepstrum coefficients</i>)
MVM	matematinės vilties maksimizavimas (angl. <i>expectation maximization</i>)

Santrumpos	
NT	neuroninis tinklas (angl. <i>neural network</i>)
NTTSP	neuroninio tinklo tarpinių svorių požymiai
AIN	asmeninis identifikavimo numeris (angl. <i>personal identification number</i>)
PMM	paslėptasis Markovo modelis (angl. <i>hidden Markov model</i>)
PS	požymių sintezė (angl. <i>feature fusion</i>)
TDA	tiesinė diskriminantinė analizė (angl. <i>linear discriminant analysis</i>)
UFM	universalus fono modelis (angl. <i>universal background model</i>)

Sąvokos	
Atpažinimas	tapatybės nustatymas pagal požymius
Autentifikavimas	varotojo tapatybės patikrinimas, parenkant priemones, kuriomis bus verifikuotas varotojo teiginys apie jo tapatybę
Identifikavimas	objekto priskyrimas artimiausiai pagal pateiktą pavyzdį objekto klasei
Verifikavimas	objekto tapatumo patikrinimas, siekiant patvirtinti objekto priskyrimą konkrečiai objekto klasei
Rekurentinis neuroninis tinklas	neuroninio tinklo tipas, kuriame yra jungtys iš vėlesniuose sluoksniuose esančių neuronų į prieš tai einančių sluoksnių neuronus

PAVEIKSLŲ SĄRAŠAS

1 pav. Diktoriaus klasifikavimo modelio apmokymo veiksmų seka.....	11
2 pav. Diktoriaus atpažinimo veiksmų seka	11
3 pav. Melų skalės kepstro koeficientų skaičiavimo seka	13
4 pav. Balso įrašo kadru ir langų sąryšis	14
5 pav. Tarpinių svorių požymių išskyrimo seka.....	16
6 pav. Maišaties matrica	19
7 pav. Pirmosios eilės Markovo grandinės atvaizdavimas	21
8 pav. Kairės-dešinės paslėptasis Markovo modelis	21
9 pav. Gauso skirstinio komponento suma	23
10 pav. Diktoriaus modelio ir fono modelių sumos diagrama	24
11 pav. Požymių sujungimo seka	28
12 pav. Pagrindinių neuroninio tinklo komponentų schema	31
13 pav. Neuroninio tinklo diagrama	32
14 pav. Žingsninės, tiesinės ir sigmoidinės funkcijos diagramos.....	34
15 pav. Tipinė dviejų sluoksnių, perdavimo pirmyn tipo, neuroninio tinklo diagrama.....	36
16 pav. Rekurentinis neuroninis tinklas su būsenos neuronais įėjimo sluoksnyje	37
17 pav. Daugiasluoksnis perceptronas.....	38
18 pav. Rekurentinio neuroninio tinklo struktūra ir informacijos sklidimas tinkle laiko žingsniu t.....	39
19 pav. Rekurentinių neuroninių tinklų užmaršumo savybė	41
20 pav. Dvikrypčio rekurentinio neuroninio tinklo struktūra.....	42
21 pav. Ilgos trumpalaikės atminties tinklo celė	43
22 pav. Neuroninis tinklas su neuronų atjungimu	46
23 pav. Principinis paslėptas Markovo modelis	51
24 pav. Principinė naudojamo neuroninio tinklo topologija	52
25 pav. Tyrimo veiksmų seka.....	53
26 pav. Melų skalės kepstro koeficientų požymių išskyrimo etapų seka.....	54
27 pav. Klasifikatorių hiper-parametrų paieškos ciklas	57

28 pav. Paslėpto Markovo modelio apmokymo veiksmų seka	58
29 pav. Rekurentinio neuroninio tinklo apmokymo veiksmų seka	59
30 pav. Detalios rezultatų analizės diagrama	61
31 pav. Siūloma neuroninio tinklo topologija	62
32 pav. Garsyno suskirstymas į bandomąją imtį ir 5 dalis pilnam eksperimentui	66
33 pav. Bandomosios imties identifikavimo tikslumo grafikas, naudojant paslėptus Markovo modelius	69
34 pav. Bandomosios imties identifikavimo tikslumo grafikas, klasifikavimui naudojant rekurentinius neuroninius tinklus.....	72
35 pav. Bandomosios imties tiksliausių rezultatų palyginimo grafikas	72
36 pav. Tiksliausių identifikavimo rezultatų palyginimas, naudojant paslėptus Markovo modelius	74
37 pav. Tiksliausių identifikavimo rezultatų palyginimas, naudojant rekurentinius neuroninius tinklus	75
38 pav. Pilnos eksperimento apimties tiksliausių identifikavimo tikslumo rezultatų palyginimas	75
39 pav. Pirmo rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą	79
40 pav. Antro rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą	79
41 pav. Trečio rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą	80
42 pav. Ketvirtą rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą	80
43 pav. Penkto rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą	81

LENTELIŲ SĄRAŠAS

1 lentelė. Biometrinių identifikatorių savybių palyginimas	8
2 lentelė. PMM, GMM, UFM klasifikavimo metodų tikslumo palyginimas diktoriaus identifikavimo uždaviniui	29
3 lentelė. DP, AVM neuroninių tinklų klasifikavimo tikslumo palyginimas diktoriaus identifikavimo uždaviniui	46
4 lentelė. Fonemų atpažinimo tikslumo rezultatų palyginimas, naudojant neuroninio tinklo klasifikatorius.....	47
5 lentelė. Klasifikatorių hiper-parametrų paieškos lentelė	50
6 lentelė. Garsyno LIEPA charakteristikos	65
7 lentelė. Garsyno LIEPA bandomojo eksperimento charakteristika.....	66
8 lentelė. Garsyno LIEPA pilno eksperimento rinkinių charakteristika.....	67
9 lentelė. PMM tinklo atpažinimo rezultatų priklausomybė nuo PMM būsenų skaičiaus	68
10 lentelė. PMM tinklo patikslinto tyrimo rezultatai su bandomąja imtimi.....	68
11 lentelė. RNT tinklo atpažinimo rezultatų priklausomumas nuo tinklo topologijos	70
12 lentelė. RNT tinklo atpažinimo rezultatai, įvertinant skirtingus atjungimo sluoksnio koeficientus.....	70
13 lentelė. RNT tinklo atpažinimo rezultatai, patikslinus atjungimo sluoksnio koeficientus	71
14 lentelė. PMM ir RNT geriausių klasifikatorių rezultatai	73
15 lentelė. PMM ir RNT geriausių klasifikatorių rezultatų vidurkiai visiems rinkiniams	74
16 lentelė. Klaidų tipo detalizacija kiekvienam tyrimo rinkiniui ir jų statistinio patikimumo koeficientai	76
17 lentelė. PMM ir RNT rezultatai naudojant kryžminio validavimo metodą.	77
18 lentelė. PMM ir RNT rezultatai tiriant aptriukšmintą signalą kryžminio validavimo metodu	78

TURINYS

IVADAS	1
1 ASMENS ATPAŽINIMO PAGAL BALSĄ METODAI	6
1.1 Biometrinio atpažinimo uždavinys	7
1.2 Asmens atpažinimas pagal balsą.....	10
1.2.1 Balso signalo požymiai	11
1.2.2 Signalų požymių išskyrimas.....	12
1.2.3 Požymių klasifikavimas	17
1.3 Klasikiniai diktoriaus identifikavimo metodai.....	20
1.3.1 Paslėptieji Markovo modeliai.....	20
1.3.2 Gauso mišinių modeliai.....	22
1.3.3 Universalus fono modelis	23
1.3.4 Jungtinių faktorių analizė	25
1.3.5 i-vektoriai	26
1.3.6 Atraminių vektorių mašinos	27
1.3.7 Požymių sintezės metodai	27
1.3.8 Klasikinių diktoriaus identifikavimo metodų tikslumas	28
1.4 Diktoriaus identifikavimas taikant neuroninius tinklus	29
1.4.1 Neuroniniai tinklai.....	30
1.4.2 Neuroninio tinklo topologijos	36
1.4.3 Apibendrinimo problema	45
1.4.4 Neuroninių tinklų diktoriaus identifikavimo metodų tikslumas ..	46
1.5 Skyriaus išvados.....	47
2 DIKTORIAUS IDENTIFIKAVIMO METODAI.....	49
2.1 Klasifikavimo metodų hiper-parametrų paieška	49

2.2	Identifikavimo tikslumo tyrimas	52
2.2.1	Požymių išskyrimas.....	53
2.2.2	Klasifikatorių mokymas ir testavimas	56
2.2.3	Rezultatų analizė ir palyginimas	60
2.3	Siūlomas diktoriaus identifikavimo metodas	61
2.4	Skyriaus išvados	63
3	EKSPERIMENTINIO TYRIMO REZULTATAI	64
3.1	Garsynas LIEPA.....	64
3.2	Tyrimai naudojant geriausią klasifikatorių	67
3.2.1	Tyrimas su bandomąja garsyno imtimi	68
3.2.2	Tyrimai naudojant pilną LIEPA garsyną.....	73
3.3	Skyriaus išvados	81
	IŠVADOS.....	83
	LITERATŪRA.....	85
	AUTORIAUS PUBLIKACIJŲ SĄRAŠAS DISERTACIJOS TEMA.....	97

IVADAS

1. Tyrimo sritis

Darbo tyrimo sritis yra asmens tapatybės identifikavimas naudojant balso įrašus. Tyrimo sritis priklauso informatikos inžinerijos mokslo sričiai, signalų technologijos mokslo šakai. Tyrimo sritis apima keleto uždavinių sprendimą: balso signalų apdorojimą, diktoriui identifikuoti tinkamų požymių radimą, diktorius balso požymių klasifikavimo uždavinį. Šiame darbe koncentruojamasi į diktorius balsą aprašančių požymių klasifikavimo problemą. Šios problemos sprendimas reikalauja daugelio skirtingų algoritmų panaudojimo. Kiekvienas algoritmas slepia savyje nemažus iššūkius: nuo balso pavyzdžių surinkimo, unikalių požymių išskyrimo iki požymių klasifikavimo ir jų priskyrimo konkrečiam diktoriui.

2. Problemos aktualumas

Asmens identifikavimas pagal balsą yra vienas iš patogiausių asmens biometrinio identifikavimo būdų. Problema yra aktuali, nes egzistuojantys diktorius identifikavimo algoritmai ir sistemos neatpažįsta diktorius 100 proc. tikslumu, tad atliekant identifikavimą visuomet lieka klasifikavimo klaidos tikimybė. Pagrindinė to priežastis yra ypač didelis signalo variatyvumas. Tai, kad kiekvieno diktorius kiekvienas ištartimas yra vis kitoks, kad balsas gali būti įtakotas ligos ar bandymų pamėgdžioti identifikuojamą asmenį, yra tik keli tokio variatyvumo pavyzdžiai. Diktorius balso nepastovumas yra pagrindinė priežastis, kodėl, nepaisant nemenkų pastangų, šioje srityje diktorius identifikavimo negalima laikyti tobulai išspręstu uždaviniu ir tema išlieka aktuali mokslinė prasme – diktorius identifikavimo tikslumą galima padidinti.

Diktorius identifikavimo tikslumas ypač aktualus, kai reikia atpažinti asmenį pagal jo balsą. Kadangi nėra sistemų, identifikuojančių diktorių 100 proc. tikslumu, jų pritaikymas yra apribotas. Pavyzdžiui, diktorius identifikavimas yra aktualus kriminalistikoje, siekiant atpažinti frazę ištarusį asmenį, tačiau neturint pakankamai aukšto tikslumo patikimumo, pritaikymas šioje srityje yra ribotas. Kitose srityse, kuriose yra leidžiama didesnė tikslumo paklaida, identifikuojant

pašnekovą galėtume pateikti daugiau personalizuotos, asmeninės informacijos, suteikti papildomų paslaugų. Anot Bimbot ir kt. (2004), diktoriaus identifikavimo uždavinys yra labai aktualus norint apsaugoti bet kokio tipo sąsajas ar autentifikuojant asmenį. Todėl bet koks diktoriaus identifikavimo tikslumo padidinimas suteiktą galimybę plačiau panaudoti identifikavimą pagal balsą minėtose srityse.

Iš kitos pusės, diktoriaus identifikavimo uždavinys taip pat yra sudėtingas dėl atpažinimui reikalingų įvairių metodų panaudojimo komplekso. Kad galima būtų sėkmingai identifikuoti diktorių, reikia surinkti jo balso pavyzdžius, išskirti šių balso pavyzdžių požymius ir atlikti klasifikavimą.

Kadangi diktoriaus identifikavimas yra klasifikavimo problema, sprendžiant klasifikavimo uždavinius gerai žinoma „Nėra nemokamų pietų“ (angl. *no free lunch*) teorema, kuri sako, kad kiekvienas klasifikatorius gali būti pakankamai tikslus tik klasifikuojant tam tikros srities specifinį uždavinį (Goodfellow, Bengio ir Courville 2016). Taigi, sprendžiamos diktoriaus identifikavimo problemos aktualumas yra dar didesnis, nes darbe eksperimentas yra atliekamas su lietuviškai kalbančių diktorių garsynu, o didelio pavyzdžių kiekio diktoriaus identifikavimo tyrimų su lietuvių kalba beveik nėra atlikta.

3. Tyrimo objektas

Disertacijos tyrimo objektas – diktoriaus identifikavimas pagal akustinį balso signalą, taikant rekurentinius neuroninius tinklus.

4. Tyrimo metodai

Disertacijos uždaviniai yra sprendžiami naudojant literatūros apžvalgos, teorinės analizės ir eksperimentinio tyrimo metodus. Teorinei analizei ir tyrimams panaudotos biometrikos sistemų, skaitmeninio signalų apdorojimo, skaitmeninio intelekto bei atpažinimo teorijos žinios.

5. Tikslas ir uždaviniai

Darbo tikslas – pasiūlyti klasifikavimo metodą, skirtą lietuviškai kalbančio diktoriaus identifikavimui pagal balsą, paremtą rekurentiniais neuroniniais tinklais.

Darbo tikslui pasiekti sprendžiami uždaviniai:

1. Ištirti neuroninių tinklų efektyvumą lietuviškai kalbančio diktoriaus identifikavimui, palyginant su kitais diktoriaus atpažinimo metodais. Įvertinti skirtingų neuroninių tinklų tipų efektyvumą diktoriaus atpažinimo uždaviniui.
2. Pasiūlyti geriausią rekurentinių neuroninių tinklų topologiją lietuviškai kalbančio diktoriaus identifikavimui.
3. Įvertinti pasiūlyto klasifikavimo metodo tikslumą skirtingomis sąlygomis, naudojant skirtingos kokybės signalus.

6. Mokslinis naujumas

Šio darbo mokslinį naujumą sudaro tokie elementai:

1. Sukurta rekurentinio neuroninio tinklo konfigūracija su 160 neuronų ir trijų paslėptų sluoksnių dvikrypčio ilgos trumpalaikės atminties topologijos konfigūracija, skirta identifikuoti lietuviškai kalbantį diktorių pagal jo balsą.
2. Parodyta, kad tokia topologija leidžia pasiekti aukščiausią diktoriaus identifikavimo tikslumą skirtingiems duomenų rinkiniams.
3. Eksperimentiniais tyrimais nustatyta, kad lietuviškai kalbančių diktorių identifikavimui naudojant dvikrypčio ilgos trumpalaikės atminties rekurentinio neuroninio tinklo klasifikatorių, gaunami nuo 3 proc. iki 6 proc. tikslesni diktoriaus identifikavimo rezultatai, nei atlikus akustinį modeliavimą su paslėptaisiais Markovo modeliais.

7. Darbo rezultatų praktinė reikšmė

Pasiūlyta neuroninio tinklo klasifikatoriaus konfigūracija gali būti naudojama siekiant pagerinti lietuviškai šnekantiems vartotojams teikiamų paslaugų kokybę ar išplėsti jų pasiūlą, teikti individulių paslaugų pasiūlymus, kai asmens tapatybę reikia nustatyti naudojant balsą.

8. Ginamieji disertacijos teiginiai

1. Lietuviškai kalbančio diktoriaus identifikavimo tikslumą galima statistiškai reikšmingai padidinti, turimam diktoriaus akustinį signalą aprašančių požymių rinkiniui parenkant tinkamesnį klasifikatorių.

2. Neuroninio tinklo klasifikatoriaus topologija, naudojant dvikryptės ilgos trumpalaikės atminties rekurentinį neuroninį tinklą, lietuviškai kalbančius diktorius identifikuoja tiksliau, nei klasifikuojant diktorius, naudojant paslėptuosius Markovo modelius.
3. Nustačius dvikrypčio ilgos trumpalaikės atminties rekurentinio neuroninio tinklo topologiją, naudojant nedidelį sluoksnių skaičių, sukuria stebimas geriausias lietuviškai kalbančio diktoriaus identifikavimo tikslumas, ir toliau keičiant tik paslėptų neuronų sluoksnių gylį, bet paliekant panašų bendrą neuronų skaičių, gaunamas tik nežymiai besiskiriantis identifikavimo tikslumas.

9. Darbo rezultatų aprobavimas

Disertacijos rezultatai buvo paskelbti 2 periodiniuose ir 6 kituose leidiniuose.

Disertacijos rezultatai pristatyti šiose tarptautinėse konferencijose:

1. BIS 2017 „20th International Conference on Business Information Systems“. Pranešimas: „Identifying Lithuanian Native Speakers Using Voice Recognition“. 2017-06-29, Poznanė, Lenkija.
2. BIS 2016 „19th International Conference on Business Information Systems“. Pranešimas: „Speaker Authentication System Based on Voice Biometrics and Speech Recognition“. 2016-07-07, Leipcigas, Vokietija.

Skaityti pranešimai respublikinėse konferencijose:

1. International Conference „Electrical, Electronic and Information Sciences“ 2018. Pranešimas: „Diktoriaus identifikavimas naudojant BLSTM tipo neuroninio tinklo konfigūraciją“. 2018-04-26, Vilnius, Lietuva.
2. 9th International Workshop „Data analysis methods for software systems“. Pranešimas: „Speaker Recognition Using Deep Neural Networks“. 2017-11-30, Druskininkai, Lietuva.

3. 8th International Workshop „Data analysis methods for software systems“. Pranešimas: „Speaker Recognition Using Deep Neural Networks“. 2016-12-01, Druskininkai, Lietuva.
4. 7th International Workshop „Data analysis methods for software systems“. Pranešimas: „GMM-UBM Enhancement with DNNs for Automated Speaker Recognition Systems“. 2015-12-03, Druskininkai, Lietuva.
5. Informacinės technologijos 2015. 20-oji tarpuniversitetinė magistrantų ir doktorantų konferencija. Pranešimas: „Asmens balso panaudojimas autentifikavimui“. 2015-04-24, Kaunas, Lietuva.

10. Disertacijos struktūra

Disertacijos apimtis 99 lapai, 18 lentelių, 43 paveikslai. Disertacijoje remtasi 92 literatūros šaltiniais. Disertaciją sudaro įvadas, 3 skyriai: Asmens atpažinimo pagal balsą metodai, Diktoriaus identifikavimo algoritmai, Eksperimentinio tyrimo rezultatai, išvados ir literatūros sąrašas.

1 ASMENS ATPAŽINIMO PAGAL BALSĄ METODAI

Šiame skyriuje yra aprašomi biometrinio atpažinimo tipai, pateikiami atpažinimo tikslumo įvertinimo kriterijai. Detaliau yra analizuojamas asmens atpažinimo pagal balsą uždavinys, kuris susideda iš požymių išskyrimo, klasifikatorių mokymo etapų. Tolesnėse skyriaus dalyse atliekama klasikinių diktoriaus identifikavimo metodų apžvalga: paslėptųjų Markovo modelių (Duda, Hart ir Stork 2000), Gauso mišinių modelių (Reynolds, Quatieri ir Dunn 2000; Ding, Yen ir Ou 2014), atraminių vektorių mašinų (Cortes ir Vapnik 1995). Toliau yra nagrinėjamas neuroninių tinklų pritaikymas diktoriaus akustiniam modeliavimui (Lei ir kt. 2014; Schmidhuber 2014). Pristatomas neuroninis tinklas (Krose ir Smagt 1996; Duda, Hart ir Stork 2000; Kamarauskas 2009), jo komponentai: aktyvavimo funkcijos (Krose ir Smagt 1996; Duda, Hart ir Stork 2000), tinklo mokymo niuansai (Richardson, Reynolds ir Dehak 2015; Keskar ir kt. 2017). Analizuojamos skirtingos neuroninių tinklų topologijos: daugiasluoksniai perceptronai (Krose ir Smagt 1996), rekurentiniai neuroniniai tinklai (angl. *recurrent neural networks*) (Schuster ir Paliwal 1997; Graves, Mohamed ir Hinton 2013; Gal ir Ghahramani 2016), ilgos trumpalaikės atminties (angl. *long short-term memory*) tinklai (Hochreiter ir Schmidhuber 1997; Schmidhuber 2014, Graves, Mohamed ir Hinton 2013). Apžvelgiama neuroninių tinklų persimokymo problema (Duda, Hart ir Stork 2000; Srivastava ir kt. 2014; Gal ir Ghahramani 2016), pristatomi kitų autorių atlikti tyrimai, taikant neuroninius tinklus diktoriaus identifikavimo (Moreno ir Ho 2003; Chen ir Luo 2009; Islam, Khan ir Haque 2013; Nijhawan ir Soni 2014; Mohanty ir Swain 2014; Ding, Yen ir Ou 2014; Nath ir Kalita 2014; Chauhan 2017; Ge ir kt. 2017) ir fonemų identifikavimo (Graves ir Schmidhuber 2005) srityse.

Skyriuje pristatoma apžvalga buvo publikuota konferencijose skaitytų pranešimų medžiagoje (Dovydaitis ir Rudžionis 2015; Dovydaitis, Rasmus ir Rudžionis 2017).

1.1 Biometrinio atpažinimo uždavinys

Biometriniam atpažinimui naudojami biologiniai matavimai, tokie kaip piršto antspaudas ar akies rainelė, kurie yra individualūs kiekvienam asmeniui. Pagrindiniai skaitmeninėse sistemose naudojami biometrinės informacijos šaltiniai yra (Jain, Dass ir Nandakumar 2004; Jain, Ross ir Prabhakar 2004; Robinson 2011; Anderson 2008):

1. Ranka rašytas parašas – labai populiarius, tačiau jį lengva suklastoti, nes retai tas pats asmuo kelis kartus pasirašo visiškai identiškai.
2. Veido atpažinimas – dažniausiai sutinkamas ir naudojamas ten, kur reikalingi asmens dokumentai, gana didelė neteisingo atpažinimo paklaida (kai kuriais atvejais – iki 20 proc.).
3. Pirštų antspaudai – užima didžiausią biometrijos rinkos dalį; patikimas sprendimas, tačiau reikalauja investicijų į įrangą; naudojamas tiek autentifikavimo uždaviniams, tiek teisminei ar kriminalistinei ekspertizei.
4. Akies rainelė – vienas iš tiksliausių biometrinių duomenų; įrangos kaina daug didesnė, nei pirštų atspaudų atpažinimo; nepatogu tai, kad akis turi būti pritraukta arti prie nuskaitymo įtaiso.
5. Balso atpažinimas – nelengva pasiekti aukštą atpažinimo tikslumą dėl didelio signalo nepastovumo; palyginti pigi įranga.

Kai kurie šaltiniai (Jain, Dass ir Nandakumar 2004; Robinson 2011) išskiria ir kitus, tačiau skaitmeniniame atpažinime rečiau sutinkamus biometrinio autentifikavimo tipus: DNR atpažinimas, identifikavimas pagal ausies formą, eiseną, rankos geometriją, kvapą, venų išsidėstymą. Šie metodai dažniausiai yra labai brangūs.

Kiekvienas iš skirtingų biometrinio autentifikavimo tipų turi savo trūkumų ir privalumų. Pavyzdžiui, kai kurie parametrai kinta su laiku – žmogui senstant. Šaltiniuose (Jain, Ross ir Prabhakar 2004; Anderson 2008; Kinnunen ir Li 2009) yra išskiriamos panašios biometrinių identifikatorių savybės, šių šaltinių apibendrinimas pateikiamas 1-oje lentelėje.

1 lentelė. Biometrinių identifikatorių savybių palyginimas

Jain, Ross ir Prabhakar (2004); Anderson (2008); Kinnunen ir Li (2009)

Požymio pavadinimas	Piršto antspaudas	Akies rainelė	Veidas	Balsas
Populiarumas	Didelis	Mažas	Vidutinis	Mažas
Atkartojimas	Sunkus	Sunkus	Vidutinis	Vidutinis
Paklaida	Drėgnumas, nešvarumai	Prastas apšvietimas	Apšvietimas, plaukai, akiniai	Triukšmas, fiziologiniai pokyčiai, mikrofono tipas
Tikslumas	Aukštas	Labai aukštas	Vidutinis	Vidutinis
Patogumas naudoti	Geras	Vidutinis	Geras	Labai geras
Užtikrinamas saugumo lygis	Aukštas	Labai aukštas	Vidutinis	Vidutinis
Požymių kitimas bėgant laikui	Mažas	Mažas	Vidutinis	Vidutinis
Atpažinimo įrenginio kaina	Vidutinė	Didelė	Maža	Maža

Idealios biometrinės savybės atitiktų tokius parametrus (Kinnunen ir Li 2009): turėtų būti labai skirtingos tarp skirtingų asmenų, turėti mažą to paties asmens parametrų kintamumą, būti atsparios triukšmui ir iškraipymams, lengvai pamatuojamos iš signalo, sunkiai pamėgdžiojamos, nepriklausyti nuo amžiaus ar kalbėtojo variacijų. Tačiau tokie biometriniai požymiai praktiškai neegzistuoja (išskyrus DNR).

Norint nustatyti, ar vartotojas gali prieiti prie tam tikros informacijos, būtina nustatyti (t. y. atpažinti) vartotojo asmens tapatybę. Kaip rašoma enciklopediniame kompiuterijos žodyne (Dagienė, Grigas ir Jevsikova 2014),

atpažinimas apibrėžiamas kaip tapatybės nustatymas pagal tam tikrą požymių rinkinį.

Autentifikavimas atliekamas naudojant tam tikrą turimą informaciją apie vartotoją. Tokios informacijos tipai sąlyginai apibrėžiami kaip „kas aš esu“, „ką žinau“ arba „ką turiu“. Siekiant užtikrinti papildomą tapatybės apsaugojimą, „ką žinau“ ir „ką turiu“ metodai gali būti sujungti, taip sukuriant dviejų tipų autentifikavimo mechanizmus. Šie du mechanizmai sujungia žiniomis grindžiamą (angl. *knowledge-based*) autentifikavimą, pavyzdžiui, slaptažodį ar AIN (asmeninis identifikavimo numeris, angl. *personal identification number*) kodą, ir papildomą identifikatorių, kuris gali būti fizinis įrenginys, pavyzdžiui, vienkartinį kodų generatorius, išmanioji kortelė (angl. *smartcard*) arba biometrinė vartotojo informacija, pavyzdžiui, akies rainelė, piršto antspaudas ar balsu ištarta frazė. Vartotojams autentifikavimą dažnai patogiau atlikti naudojant biometrinę informaciją, kadangi vartotojas ją visada „turi su savimi“ ir to nereikia įsiminti.

Identifikavimas (objekto priskyrimas artimiausiai pagal pateiktą pavyzdį objekto klasei) ir verifikavimas (tapatumo patikrinimas) yra du skirtingi uždaviniai, kurie iš pirmo žvilgsnio gali pasirodyti labai panašūs.

Identifikavimas yra pateikto pavyzdžio palyginimas ir išskyrimas iš visų turimų pavyzdžių. Tarkime, norime identifikuoti asmenį pagal ištartą frazę, tad atliekame paiešką visoje turimoje ištartimų duomenų bazėje ir palyginę rezultatus teigiame, jog diktorius yra X asmuo (Naik 1990; Kamarauskas 2009). Šio uždavinio sprendimas yra klasifikacijos (objektų skirstymo į grupes pagal tam tikrus požymius) rezultatas, priskiriant pateiktą pavyzdį vienai iš pavyzdžių grupių.

Identifikavimo sistemos gali būti dviejų tipų – sistemos, kurios būtinai priima sprendimą, parenkant panašiausią į pateiktą pavyzdį, ir sistemos, kurioms yra pritaikoma tikslumo tolerancijos riba. Pasiekus šią ribą, priimamas sprendimas, kad diktorius nepriklauso turimai pavyzdžių aibei, t. y. sistema iš esmės nepriima jokio sprendimo apie asmens tapatybę.

Verifikavimas atliekamas duomenų bazėje, kurioje saugomi tik ieškomo diktoriaus pavyzdžiai. Tarkime, diktorius pasako frazę, teigdamas, jog jis yra diktorius X. Toliau atliekame paiešką tarp visų turimų diktoriaus X pavyzdžių ir nustatome, ar diktorius X yra tikrai tas, kuris sakosi esantis (Naik 1990).

Verifikavimo atveju, pagal nustatytą tolerancijos ribą priimame sprendimą, ar pateiktas pavyzdys pakankamai tiksliai atitinka turimą diktoriaus pavyzdžių aibę.

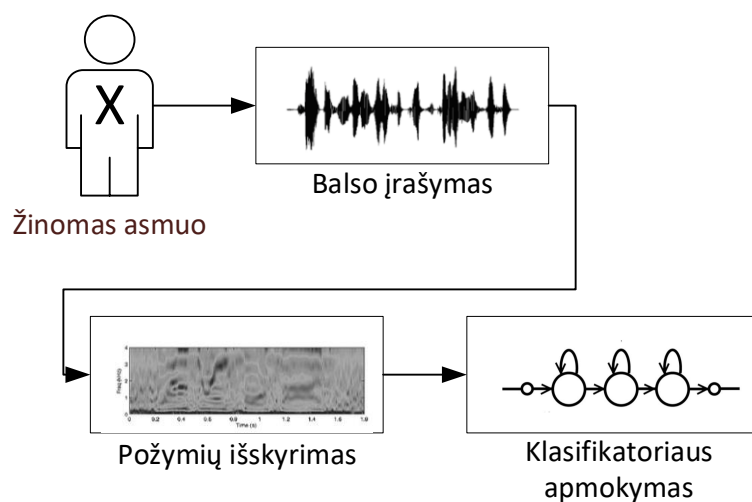
Diktoriaus klasifikacijos uždavinys taip pat priklauso nuo to, ar rezultatas būtinai turi būti priskirtas konkrečiam asmeniui. Tiek identifikavimas, tiek verifikavimas gali būti priklausomas arba nepriklausomas nuo pasakyto teksto (Šalna ir Kamarauskas 2010). Šis aspektas yra svarbus tuo požiūriu, jog nuo jo priklauso parinkto klasifikatoriaus pritaikomumas.

1.2 Asmens atpažinimas pagal balsą

Asmens atpažinimas pagal balsą iš kitų biometrinės informacijos tipų išsiskiria tuo, jog jam nereikalinga brangi įranga vartotojo pusėje. Be to, asmens atpažinimas pagal balsą yra patogus būdas perduoti biometrinę informaciją per atstumą. Kita vertus, balso parametrai kinta, tiek priklausomai nuo žmogaus emocinės būsenos, galimų susirgimų, nuo įrangos, kuri naudojama informacijos įrašui atlikti, tiek nuo amžiaus.

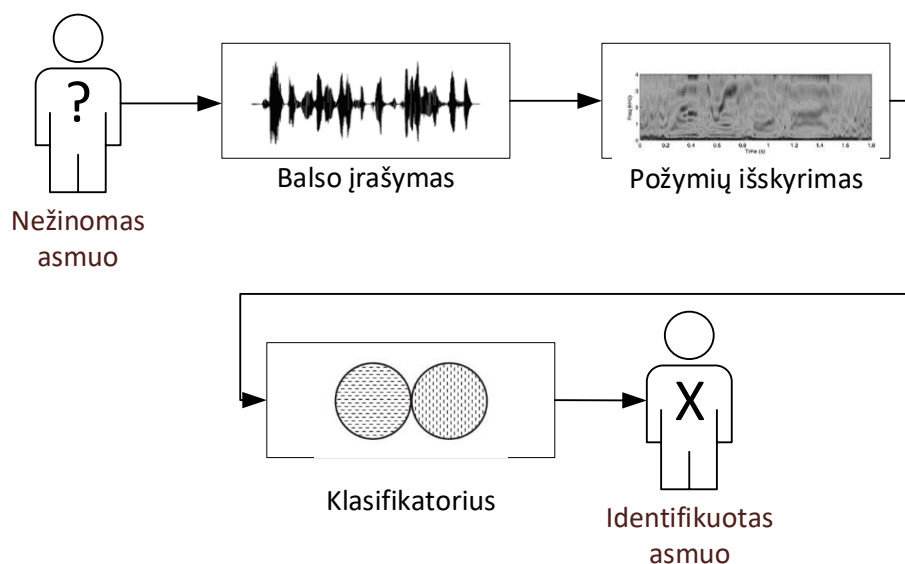
Kad galėtume atlikti diktoriaus atpažinimą pagal balsą, turime įvykdyti du skirtingus etapus (Šalna ir Kamarauskas 2010). Pirmojo, diktoriaus balso pavyzdžių kaupimo, metu yra vykdoma veiksmų seka, parodyta 1-ame paveiksle.

Šiame paveiksle matome, kad iš anksto žinomo diktoriaus balso pavyzdys yra įrašomas į skaitmeninę laikmeną. Tuomet iš turimo balso pavyzdžio turi būti išgauti signalo požymiai, kurie yra maksimaliai unikalūs kiekvienam kalbėtojų. Pagal šiuos požymius yra atliekamas klasifikatoriaus apmokymas.



1 pav. Diktoriaus klasifikavimo modelio apmokymo veiksmų seka

Diktoriaus atpažinimas yra vykdomas pagal veiksmų seką, pavaizduotą 2-ame paveiksle.



2 pav. Diktoriaus atpažinimo veiksmų seka

Šiame paveiksle matome, kaip imamas nežinomo diktoriaus balso įrašas, iš jo išskiriami tie patys požymiai, kaip mokymo metu. Vėliau požymiai yra perduodami į apmokytą klasifikatorių, kurio pagalba konkretaus balso pavyzdžio požymiai yra priskiriami konkrečiai klasei, kuri atitinka konkretų diktorių.

1.2.1 Balso signalo požymiai

Išanalizavus mokslinės literatūros šaltinius, galima apibendrinti, kad pagrindiniai požymių išskyrimo metodai naudoja įvairias dažnių transformacijas

(Togneri ir Pullella 2011). Kituose tyrimuose (Lei ir kt. 2014), ypač šnekamosios kalbos atpažinimo srityje, yra pažymimos neuroninių tinklų galimybės. Juos aptarsime 1.2.2 skyriuje.

Naudojami kalbos signalų požymiai gali būti apibendrinti kaip (Kinnunen ir Li 2009): trumpalaikės spektrinės savybės; balso šaltinio savybės; spektro dinamikos savybės; fonetinės savybės; aukšto lygio savybės (individualus asmens naudojamas žodynas). Naudotinus požymius reikia pasirinkti priklausomai nuo sprendžiamo uždavinio, turimų skaičiavimo pajėgumų bei kitų parametrų. Tobulas požymis būtų toks, kuris (Kinnunen ir Li 2009):

- Stipriai skiriasi tarp skirtingų diktorių
- Neįtakojamas arba mažai įtakojamas foninio triukšmo ir iškraipymų
- Dažnai sutinkamas natūralioje kalboje
- Yra lengvai pamatuojamas
- Yra sunkiai pamėgdžiojamas
- Neįtakojamas diktorius amžiaus ar ilgalaikių variacijų diktorius balse

1.2.2 Signalo požymių išskyrimas

Signalo požymių išskyrimui naudojami šie metodai (Kinnunen ir Li 2009):

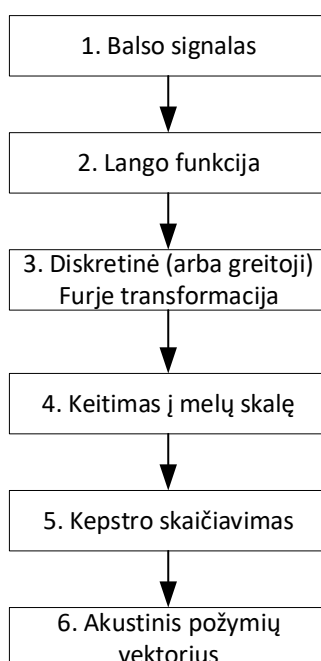
- Tiesinės prognozės keptriniai koeficientai (angl. *linear predictive cepstral coefficients*)
- Tiesiniai spektriniai dažniai (angl. *linear spectral frequencies*)
- Percepcinės tiesinės prognozės koeficientai (angl. *perceptual linear prediction coefficients*)
- Melų skalės keptrų koeficientai (angl. *mel frequency cepstral coefficients*)

Kiti naudojami požymiai (Naik 1990; Lipeika 2005; Kinnunen ir Li 2009):

- Daliniai koreliacijos koeficientai (angl. *partial correlation coefficients*);
- Logaritminių plotų tankiai (angl. *log area ratios*);
- Formančių dažniai bei pločiai (angl. *formant frequencies and bandwidths*).

Kaip teigia A. L. Lipeika (2005), B. Šalna ir J. Kamarauskas (2010), K. Kumar, C. Kim ir R. M. Stern (2011), R. Togneri ir D. Pullella (2011), šiuo metu kalbos atpažinimui dažniausiai naudojami kalbos suvokimu grįsti melų skalės keptriniai koeficientai (MSKK) su MSKK požymių pirmos ir antros eilės išvestinėmis (Šalna ir Kamarauskas 2010; Kumar, Kim ir Stern 2011). Taip pat naujesniuose tyrimuose (Variani ir kt. 2014) yra vertinamas požymių išskyrimas naudojant neuroninius tinklus. H. Arsikere, H. A. Gupta ir A. Alwan (2014), E. Variani ir kt. (2014) parodo atpažinimo tikslumo pagerinimą, naudojant požymių sujungimo metodą.

Melų skalės keptro koeficientai. Melų skalė – kalbos suvokimu grįsta netiesinė dažnių skalė. Šių koeficientų skaičiavimas (Molau ir kt. 2001; Aroon ir Dhonde 2015; Rubal ir Mehan 2017) gali būti apibendrintas veiksmų seka, pateikta 3-ame paveiksle.



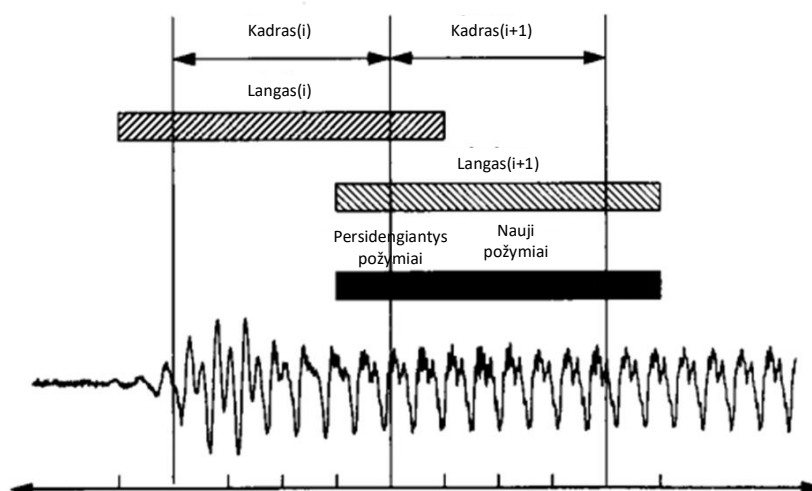
3 pav. Melų skalės keptro koeficientų skaičiavimo seka

Kepstriniai požymiai nusako signalo spektro vaizdavimą logaritminiu masteliu, o melų skalės keptro rinkinys sudaro melų skalės keptro koeficientus. Dažnių skalė melais yra susieta su tiesine dažnių skale tokia priklausomybe (Zheng, Zhang ir Song 2001):

$$M(f) = 2595 * \log_{10} \left(1 + \frac{f}{700}\right), \quad (1)$$

kur f yra signalo dažnis.

Pirmiausia yra paimamas skaitmenizuotas balso signalas. Signalui gali būti atlikta pirminė filtracija, siekiant sumažinti triukšmą. Antras žingsnis – balso signalo įrašas yra dalinamas į mažesnės trukmės atkarpas – kadrus (angl. *frames*). Kiekvienam kadrui yra pritaikoma lango funkcija. Dažnai naudojamas gretimų kadrų tarpusavio persidengimas. Tai yra iliustruojama 4-ame paveiksle.



4 pav. Balso įrašo kadrų ir langų sąryšis

Picone (1993)

Kaip išskiria W. J. Picone (1993) ir J. Kamarauskas (2009), sutinkami šie langų tipai: stačiakampis, Hamming, Hanning, Blackman, Bartlett ir Kaiser. Kalbos atpažinime beveik išskirtinai naudojamas Hammingo langas, kurio funkcija yra:

$$w(n) = 0,54 - 0,46\cos\left(\frac{2\pi n}{N_s-1}\right), \quad (2)$$

kur n yra diskretinis laikas, $0 \leq n < N_s$, $w(n) = 0$ kitais atvejais. N_s – lango trukmė imčių skaičiumi.

Sekantis, trečias žingsnis yra balso signalo diskretų transformacija į dažnines dedamąsias, tai atliekama naudojant diskretinę Furjė transformaciją (DFT) (Zheng, Zhang ir Song 2001). Praktikoje vietoje DFT beveik visada taikoma greitoji Furjė transformacija (GFT) (Deshmukh ir Bachute 2013).

Balso įrašo dažnių dedamoji yra keičiama į melų skalės gaubtinę, pagal (1)-ą formulę. Kiekvienai spektro dažnio gaubtinei melų skalėje, taikant diskretinę kosinusinę transformaciją (DKT), apskaičiuojamas melų skalės kepstras. Tokios transformacijos rezultatas yra vadinamas melų skalės kepstro koeficientu. Kadangi DKT atliekamas visoms melų skalės spektro dažnio gaubtinėms, šių koeficientų rinkinys vadinamas MSKK, kuris taip pat vadinamas akustinių požymių vektoriumi (Muda, Begam ir Elamvazuthi 2010).

Kaip teigia autoriai (Tiwari 2010; Ravindran, Anderson ir Slaney 2006), melų skalės kepstro koeficientai yra vienas atspariausių būdų, aprašant laike kintančias savybes.

Pirmos ir antros eilės melų skalės kepstro koeficientų išvestinės. Dažniausiai kartu su MSKK yra naudojami pirmos eilės ir antros eilės išvestinių koeficientai, kurie taip pat yra vadinami delta bei delta-delta arba akceleracijos koeficientais (Ringelienė ir Filipovič 2011; Šalna ir Kamarauskas 2010; Kumar, Kim ir Stern 2011; Togneri ir Pullella 2011).

Delta ir delta-delta koeficientai yra apskaičiuojami pagal formulę (Young ir kt. 2005; Togneri ir Pullella 2011):

$$\delta_t = \frac{\sum_{p=1}^P p(c_{t+p} - c_{t-p})}{2 \sum_{p=1}^P p^2}, \quad (3)$$

kur δ_t yra delta koeficientas laiko momentu t , apskaičiuotas MSKK koeficientams nuo $c_{t-\theta}$ iki $c_{t+\theta}$, P yra delta lango dydis.

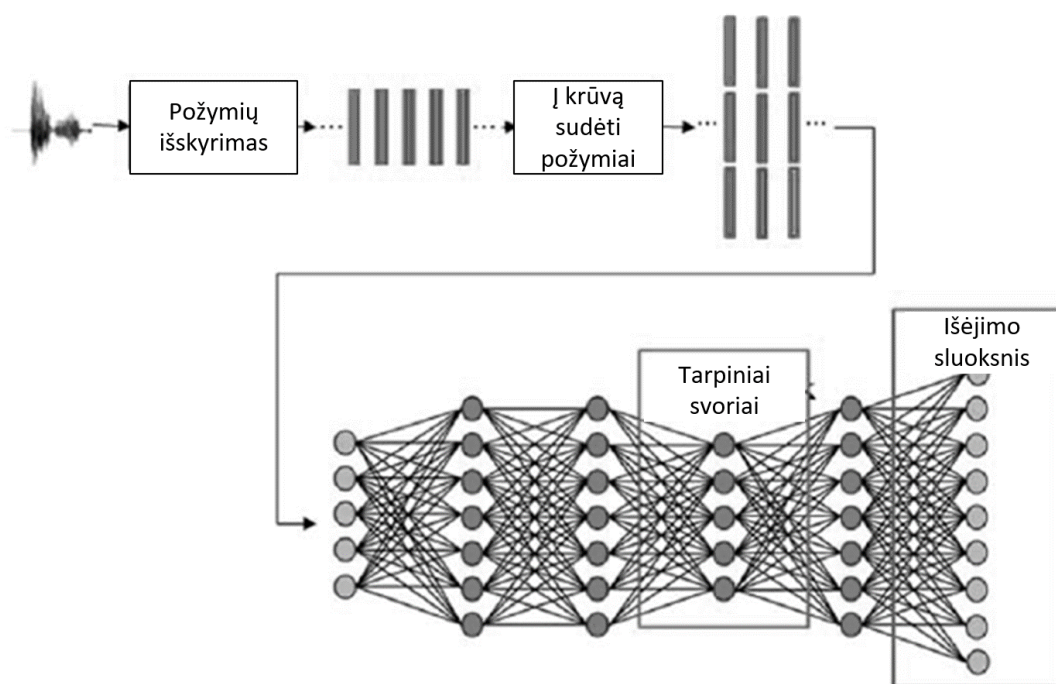
R. Togneri ir B. Pullella (2011) pažymi, jog naudojant minėtas išvestines yra išgaunama informacija apie diktoriaus šnekos stilių ir trukmę, bei pateikiamas 39 požymių rinkinio pavyzdys, kuris yra sudaromas iš 13 MSKK, 13 delta, bei 13 delta-delta koeficientų. Jis yra skaičiuojamas kiekvienam balso kadrai ir atrodo taip (Šalna ir Kamarauskas 2010; Togneri ir Pullella 2011):

- 12 MSKK koeficientų;
- 1 MSKK kadro energijos elementas;
- 12 delta – koeficientų;
- 1 delta kadro energijos elementas;
- 12 delta-delta – koeficientų;

- 1 delta-delta kadro energijos elementas.

Kitos požymių išskyrimo ir sujungimo technikos. Kaip minėta 1.2.2 skyriaus pradžioje, tarp naujesnių požymių išskyrimo metodų (Variani ir kt. 2014; Arsikere, Gupta ir Alwan 2014; Richardson, Reynolds ir Dehak 2015; Himawan ir kt. 2015) galime priskirti standartinį požymių rinkinį papildančius metodus. Keletas tokių yra neuroninio tinklo tarpinių svorių požymiai bei požymių sujungimo metodų taikymas.

Neuroninio tinklo tarpinių svorių požymiai. Neuroninio tinklo tarpinių svorių požymių (NTTSP) metode (Richardson, Reynolds ir Dehak 2015) yra atliekami 3-ame paveiksle išvardinti požymių išskyrimo etapai. Toliau gauti požymių vektoriai yra perduodami į neuroninį tinklą, kaip įėjimo reikšmės. Šio neuroninio tinklo klasifikavimo tikslas yra parinkti teisingą klasę, pagal pateiktą požymių vektorių. Metodas yra iliustruojamas 5-ame paveiksle.



5 pav. Tarpinių svorių požymių išskyrimo seka
Richardson, Reynolds ir Dehak (2015)

Metodas parodo iš balso signalo išskirtų požymių sudėjimą į krūvą (angl. stack) ir paruošimą neuroninio tinklo įėjimams. Neuroninis tinklas yra

apmokomas kaip įprasta, tačiau kaip šio metodo išėjimai yra paimami tinklo tarpiniai svoriai.

Autoriai F. Richardson, D. Reynolds ir N. Dehak (2015) siūlo, kad šis požymių išskyrimo metodas būtų naudojamas perduodant požymius antriniam klasifikatoriui, pavyzdžiui, perduodant į kitą gilų neuroninį tinklą. Autorių (Richardson, Reynolds ir Dehak 2015) teigimu, kai kuriems klasifikatoriams daugiamačiai vektoriai yra per dideli, todėl požymių sumažinimas vykdomas naudojant tiesinę analizę. Neuroninio tinklo tarpinių svorių požymių išskyrimo metodas gali būti pritaikytas tiek kalbos, tiek šnekos atpažinime, jei šiuo metodu apskaičiuotus požymius perduotume į klasifikatorių.

E. Variani ir kt. (2014) savo tyrime parodo, kad atliekant nedidelės apimties diktoriaus verifikavimo uždavinį, neuroninis tinklas, apmokytas naudojant neuroninio tinklo tarpinių svorių požymių išskyrimo metodą, tikslumu nusileidžia MSKK metodui. Autoriai pažymi, kad tikslumo pagerinimas stebimas, kai gaunamas bendras įverčio rezultatas su MSKK ir NTTSP požymiais, naudojant požymių sintezės metodą, kuris plačiau aprašomas 1.3.7 skyriuje.

1.2.3 Požymių klasifikavimas

Kaip teigiama 1.1 skyriuje, norime, kad asmens balso biometriniai požymiai būtų kuo labiau atsiskyrę nuo kitų diktorių tų pačių balso požymių. Individualūs požymių rinkiniai dažnai suformuoja grupes, kurios sudaro klasterius.

Klasifikatorių mokymas. Siekiant atskirti dvi ir daugiau klasių tarpusavyje, yra naudojami klasifikatoriai. Klasifikatoriai pradžioje turi būti apmokomi. Anot šaltinių (Duda, Hart ir Stork 2000; Shavlik, Dietterich ir Dietterich 1990; Russell ir Norvig 2009, Chapelle, Scholkopf ir Zien, 2006), pagrindiniai mokymo būdai yra:

1. Mokymas su mokytoju (angl. *supervised learning*) – klasifikatoriaus mokymo būdas, kai „mokytojas“ kiekvienam pavyzdžiui priskiria kategoriją. Klasifikatorius pagal pavyzdžius su priskirtomis kategorijomis

apsimoko atlikti klasifikavimą. Šio tipo klasifikatorių pavyzdžiai yra Bayes naivusis klasifikatorius, logistinė regresija, sprendimų medžiai, neuroniniai tinklai (NT).

2. Mokymas be mokytojo (angl. *unsupervised learning*) – tai toks klasifikatoriaus apsimokymo tipas, kai yra pateikiami požymiai, o pats klasifikatorius sukuria atitinkamas grupes pagal pateiktus įvesties duomenis. Šio tipo klasifikavimo algoritmų pavyzdžiai: k-vidurkių klasifikavimas, mišinių modeliai, neuroniniai tinklai ir kt.
3. Dalinis mokymas su mokytoju (angl. *semi-supervised learning*) – tarpinis mokymo variantas tarp mokymo su mokytoju ir mokymo be mokytojo. Algoritmui yra pateikiami nesužymėti mokymo pavyzdžiai su daline žymių informacija, tačiau ne visiems pavyzdžiams.
4. Pastiprintas mokymas (angl. *reinforcement learning*) – tai mokymo tipas, kai klasifikatorius spėja veiksmo priskyrimą atitinkamai klasei, dėl tam tikro paskatinimo. Toliau mokytojas arba „kritikas“ pažymi, ar klasifikatoriaus priimtas dvejetainas sprendimas buvo tinkamas, ar ne. Šio tipo mokymas yra naudingas, kai atspėti galutinį rezultatą yra per sudėtinga dėl sudėtingos veiksmų sekos. Tokio mokymo pavyzdys – Markovo sprendimų procesai (angl. *Markov decision processes*).

Tinkamiausio mokymo tipo pasirinkimas pirmiausia priklauso nuo sprendžiamo uždavinio, apmokymui skirtų duomenų kiekio ir kitų susijusių faktorių.

Klasifikavimo klaidų ir tikslumo įvertinimo matai. Norint įvertinti atpažinimo tikslumą, turime jį pamatuoti. Kadangi, sprendžiant skirtingus atpažinimo uždavinius, „klaidos“ kaina yra skirtinga, atpažinimo sistemų tikslumą tam tikrais atvejais naudinga matuoti įvertinant klaidą, ją skaičiuojant pagal skirtingus kriterijus.

Moksliniuose šaltiniuose (Jain, Dass ir Nandakumar 2004; Jain, Ross ir Prabhakar 2006; Kamarauskas 2009) teigiama, jog biometrinės informacijos atpažinimo tikslumas įvertinamas skaičiuojant autentifikavimo klaidų skaičių kaip dalį nuo visų autentifikuotų asmenų kiekio. Kadangi klaidų kaina ne visada

yra vienoda, autentifikavimo klaidos dažnai skirstomos į klaidingo priėmimo ir klaidingo atmetimo klaidas.

Statistinės binarinių klasifikatorių pirmos ir antros rūšies klaidos ir jų vertinimo matai gaunami iš maišaties matricos (angl. *confusion matrix*), pavaizduotos 6 paveiksle.

		Tikroji klasė	
		Priėmimas	Atmetimas
Spėjimo klasė	Priimti	Teisingi priėmimai (TP)	Klaidingi priėmimai (KP)
	Atmesti	Klaidingi atmetimai (KA)	Teisingi atmetimai (TA)

6 pav. Maišaties matrica
Fawcett (2005)

Klaidų matai yra naudojami, kai reikia statistiškai įvertinti identifikavimo ir verifikavimo tikslumą. Atliekant klasifikavimą su daugiau nei 2 klasėmis, binarinių klasifikatorių tikslumo įverčius galima panaudoti taikant teisingam atpažinimui teigiamą sprendimo rezultatą, o atpažinus pateiktą pavyzdį, kaip bet kurią neteisingą klasę, priskirti šią klaidą kaip neteisingą atpažinimą. Tikslumas naudojant pateiktą paveikslą būtų visų teisingų priėmimų ir teisingų atmetimų santykis su visais paveiksle paminėtais atvejais: teisingi priėmimai, klaidingi priėmimai, klaidingi atmetimai, teisingi atmetimai. Tokiu atveju tikslumas būtų apskaičiuotas taip:

$$Tikslumas = \frac{TP+TA}{TP+KP+KA+TA}, \quad (4)$$

čia TP ir TA – visų teisingų priskyrimų klasėms suma, KP ir KA – visų klaidingų priskyrimų klasėms suma.

1.3 Klasikiniai diktoriaus identifikavimo metodai

Diktoriaus atpažinimo uždaviniui naudojami klasifikavimo algoritmai priklauso nuo vykdomo atpažinimo tipo (Šalna ir Kamarauskas 2010):

- Priklausomo nuo teksto
- Nepriklausomo nuo teksto

Identifikavimas bei verifikavimas gali būti atliekami pagal vieną diktoriaus ištartą frazę ar pagal bet kokią to paties diktoriaus ištartą tekstą. Jei vykdomas nuo teksto nepriklausomas atpažinimas, naudojami Gauso mišinių modeliai (Reynolds, Quatieri ir Dunn 2000; Ding, Yen ir Ou 2014), dirbtiniai neuroniniai tinklai (Nath ir Kalita 2014; Srinivas, Rani ir Madhu 2014), atraminių vektorių mašinos (Cortes ir Vapnik 1995). Priklausomam nuo teksto atpažinimui dažniausiai naudojami paslėptieji Markovo modeliai. Pastaraisiais metais abiem uždaviniams sėkmingai pradėti taikyti giliojo mokymo neuroniniai tinklai (Garcia-Romero ir kt. 2014; Lei ir kt. 2014; Reynolds ir Dehak 2015).

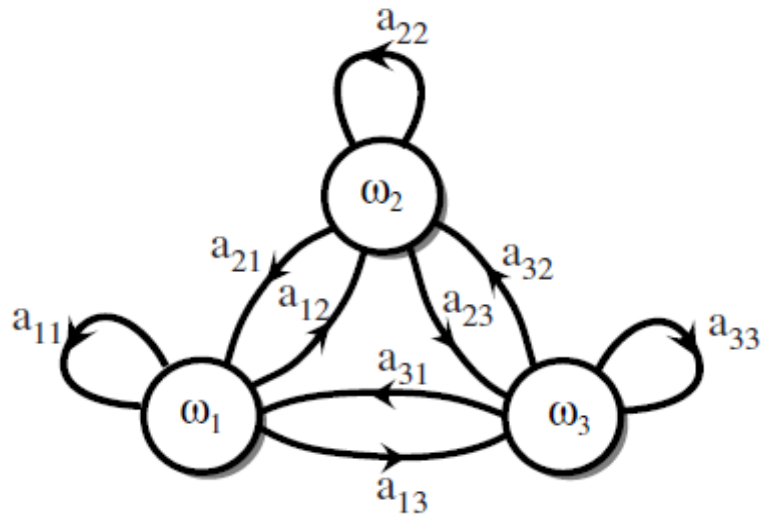
1.3.1 Paslėptieji Markovo modeliai

Kaip teigia R. O. Duda, P. E. Hart ir D. G. Stork (2000), paslėptieji Markovo modeliai (PMM) yra labai sėkmingai taikyti atliekant balso ir gestų atpažinimą. Trumpa jų charakterizacija aprašoma sekančiose pastraipose.

Būsena laike $t + 1$ gali būti įtakota prieš tai buvusios būsenos t . Būsena t bet kuriuo laiku yra pažymėta kaip $w(t)$. Ilgio T būsenų eilė yra pažymėta kaip $w^T = \{w(1), \dots, w(T)\}$. Sekos nustatymui naudojame sąlyginę tikimybę:

$$P(w_j(t)|w_i(t-1)) = a_{ji}, \quad (5)$$

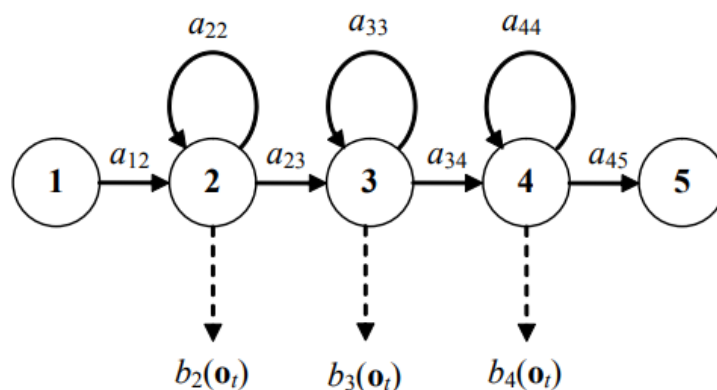
kur a_{ji} – tai nuo laiko nepriklausanti tikimybė, turinti w_j būseną t , priimant, kad laiko momentu $t - 1$ būsena buvo w_i (Duda, Hart ir Stork 2000). Šis ryšys tarp skirtingų sekų yra parodomas 7-ame paveiksle. Jame atvaizduotos diskrečiosios būsenos $w_1; w_2; w_3$ ir perėjimai tarp šių būsenų $a_{11}; a_{21}; a_{22}; a_{32}$ ir t. t.



7 pav. Pirmosios eilės Markovo grandinės atvaizdavimas
Duda, Hart ir Stork (2000)

Standartiniai PMM gali turėti ryšį tiek į pradinį, tiek į atgalinį kalbos momentą. Kaip teigia R. O. Duda, P. E. Hart ir D. G. Stork (2000), kalbos atpažinimo praktikoje beveik visi naudojami PMM yra taip vadinami „iš kairės į dešinę“ topologijos modeliai, kadangi kalbos sklidimas nebūna atgalinės krypties.

Ž. Ringelienė ir M. Filipovič (2011) kalbos atpažinimui taip pat naudoja „iš kairės į dešinę“ topologijos paslėptąjį Markovo modelį. Jis yra atvaizduojamas 8-ame paveiksle.



8 pav. Kairės-dešinės paslėptasis Markovo modelis
Ringelienė ir Filipovič (2011)

8-ame paveiksle $a_{12}, a_{22}, a_{23}, \dots, a_{45}$ yra žymimos sekos perėjimo iš 1 į 2, iš 2 į 2, ... , iš 4 į 5 tikimybės, o $b_j(o_t)$ yra stebėjimų tikimybinio tankio funkcija, priskirta j -ajai būsenai:

$$b_j(o_t) = p(o_t | s_t = j), j = 2, \dots, N - 1, \quad (6)$$

kur j – paslėpto Markovo modelio būsena, t – laiko momentas, o – stebimas požymių vektorius, s_t – būsena laiko momentu t .

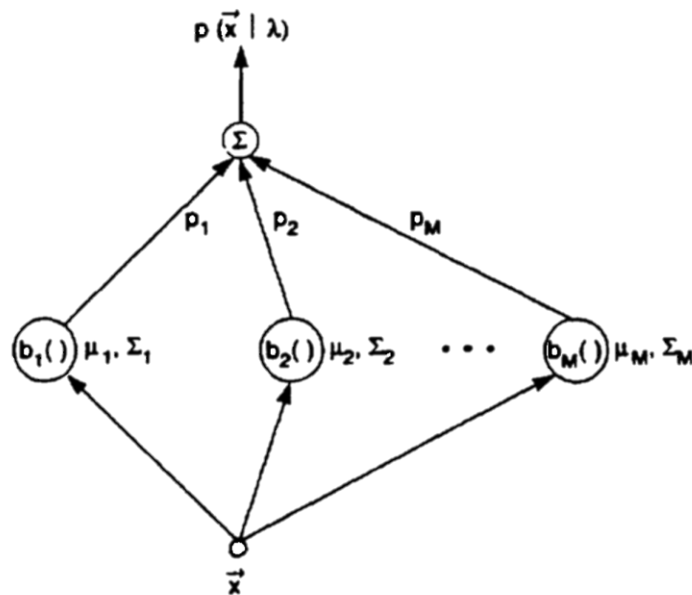
Autoriai (Fakotakis, Georgila ir Tsopanoglou 1997; Mahola, Nelwamondo ir Marwala 2007; Abdallah, Osman ir Mustafa 2012; Deshmukh ir Bachute 2013) praktiškai taiko PMM diktoriaus atpažinimo uždaviniui. Gautų tyrimų rezultatai yra apibendrinti 2-oje lentelėje, skyriaus pabaigoje.

1.3.2 Gauso mišinių modeliai

Kaip savo darbe (Reynolds, Quatieri ir Dunn 2000) teigia autoriai, Gauso mišinių modeliai (GMM) (angl. *Gaussian mixture models*) yra vienas iš dominuojančių nuo teksto nepriklausomų diktorių atpažinimo būdų. Jis dažniausiai naudojamas kaip parametrinio modelio tikimybių pasiskirstymo matavimas, pavyzdžiui, balso trakto spektrinių savybių. GMM parametrai yra skaičiuojami pagal mokymo duomenis, naudojant matematinės vilties maksimizavimo (MVM) (angl. *expectation maximization*) algoritmą. Gauso mišinio modelis yra svertinė M mišinių svorių Gauso skirstinio komponento suma, kuri gali būti užrašoma kaip:

$$p(x|\lambda) = \sum_{i=1}^M p_i b_i(x), \quad (7)$$

kur x yra D -dimensijų požymių vektorius, $p_i, i = 1, \dots, M$ yra mišinio svoriai, ir $b_i(x), i = 1, \dots, M$ yra Gauso skirstiniai. M Gauso skirstinių suma atvaizduojama 9-ame paveiksle.



9 pav. Gauso skirstinio komponento suma
Reynolds ir Rose (1995)

9-ame paveiksle pateikta M Gauso skirstinių suma, kuri yra sudaroma iš x vektoriaus, apskaičiuojant komponentinius Gauso skirstinius $b_1, b_2, \dots, b_M$, juos padauginant iš mišinio svorių $p_1, p_2, \dots, p_M$.

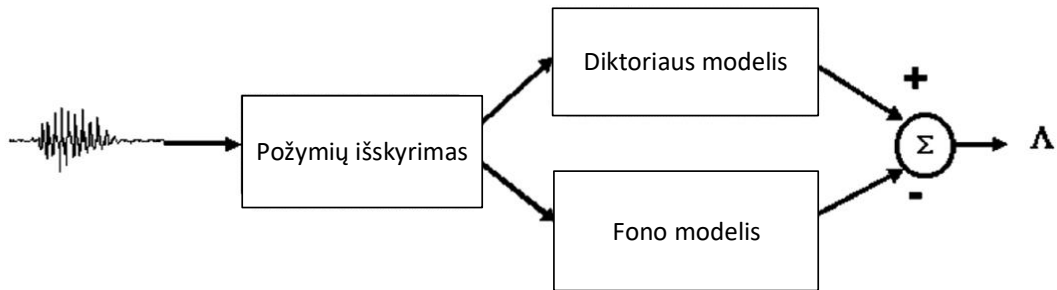
Identifikavimo rezultatų tikslumas, naudojant GMM, analizuojamas autorių darbuose (Reynolds 1995; Reynolds ir Rose 1995; Narayanaswamy 2005; Megloulis ir Khebli 2015; Aroon ir Dhonde 2015). Gauti rezultatai yra apibendrinti 2-oje lentelėje, skyriaus pabaigoje.

1.3.3 Universalus fono modelis

D. A. Reynolds, F. T. Quatieri ir R. B. Dunn (2000) teigimu, universalus fono modelis (UFM) (angl. *Universal Background Model*) yra modelis, naudojamas biometrinio verifikavimo sistemose, siekiant atvaizduoti apibendrinamąsias, nuo asmens nepriklausomas požymių charakteristikas. Šios charakteristikos, atliekant priėmimo arba atmetimo sprendimą, yra palyginamos su modeliu, kuris buvo sugeneruotas konkrečiam asmeniui.

Diktorius atpažinimo sistemose universalus fono modelis yra nuo diktorius nepriklausomas Gauso mišinys, apmokytas su kalbos pavyzdžiais iš didelės diktorių duomenų bazės. Taip siekiama atvaizduoti bendrinius kalbos bruožus. Gauso mišinių modelių mokymui naudojamas matematinės vilties

maksimizavimo algoritmas. UFM modeliai gali būti naudojami mokant nuo diktoriaus priklausomą modelį, kuris gali būti naudojamas kaip pirminis modelis maksimalios tikimybės parametrų apskaičiavime.



10 pav. Diktoriaus modelio ir fono modelių sumos diagrama
Reynolds, Quatieri ir Dunn (2000)

UFM modelis pavaizduotas 10-ame paveiksle, kuriame matome, kaip įeinantis garso signalas yra apdorojamas išskiriant požymius (1.2.2 skyrius). Kitame etape yra sukuriami du modeliai – diktoriaus modelis λ_{hyp} ir fono modelis $\lambda_{\overline{hyp}}$, kuriame turi būti visa likusi informacija. Iš šių dviejų modelių yra suskaičiuojamas logaritminis tikėtinumo santykis (Reynolds, Quatieri ir Dunn 2000):

$$\Lambda(X) = \log p(X|\lambda_{hyp}) - \log p(X|\lambda_{\overline{hyp}}), \quad (8)$$

kur $p(X|\lambda_{hyp})$ yra tikimybės tankio funkcija.

Pavyzdžiui, verifikavimo sistemoje, naudojančioje telefoninio ryšio kokybę ir tik vyriškos lyties diktorių pavyzdžius, fono modelis būtų generuojamas tik iš vyriškos lyties įrašų rinkinio. Jeigu iš anksto šie parametrai apie apmokymo pavyzdžius nėra žinomi, apmokymas yra vykdomas naudojant tiek vyriškus, tiek moteriškus pavyzdžius. Pats paprasčiausias būdas tokiam apmokymui – sujungti visus duomenis ir juos panaudoti, UFM apmokymui panaudojant MVM algoritmą.

Reikėtų atkreipti dėmesį, kad, siekiant aukšto tikslumo, mokymo pavyzdžiai turi būti tolygiai pasiskirstę tarp kategorijų, pagal kurias bus vykdomas atpažinimas. Tarkime, naudojant nuo lyties nepriklausančius duomenis, reikėtų įsitikinti, kad turime balansą tarp vyriškų ir moteriškų pavyzdžių. Priešingu atveju modelis bus šališkas dominuojančios pavyzdžių

kategorijos atžvilgiu. Tokios pačios prielaidos turi būti daromos ir įrašams iš skirtingų mikrofonų. Kitas būdas spręsti šį uždavinį – apmokyti individualius universalius modelius kiekvienai pavyzdžių kategorijai, tarkime, kiekvienai vyriškai ir moteriškai grupei, o gavus rezultatus, juos sujungti į vieną bendrą grupę. Šis būdas padidina patikimumą, naudojant nesubalansuotus duomenis, ir suteikia galimybę kontroliuoti galutinį UFM modelį.

UFM tikslumą diktorius atpažinimo uždaviniuose tyrinėja R. Zheng, S. Ulang ir B. Xu (2004); S. Cheng, I. F. Chen ir H. M. Wang (2010); K. M. Omar ir J. Pelecanos (2010); F. Ganjeizadeh ir kt. (2014); A. S. Bawaskar ir P. N. Kota (2015); M. Jayanth ir R. B. Roja (2016). Šiuose tyrimuose gauti rezultatai apibendrinami 2 lentelėje, skyriaus pabaigoje.

1.3.4 Jungtinių faktorių analizė

Jungtinių faktorių analizės (JFA) (angl. *joint factor analysis*) modelis apibrėžiamas kaip Gauso skirstinio PMM super-vektoriai (angl. *super-vector*), priklausantys nuo diktorius ir įrašo veiksnių. Super-vektorius – tai į krūvą sudėti (angl. *stacked*) nuo diktorius priklausantis vektorius ir nuo sesijos priklausantis vektorius (Kenny 2006).

Nuo diktorius ir įrašo priklausantis super-vektorius apibrėžiamas kaip

$$M = s + c, \quad (9)$$

kur s yra nuo diktorius priklausomas super-vektorius, c yra nuo sesijos priklausomas super-vektorius.

Diktorius veiksniai bei įrašo veiksniai atlieka skirtingus vaidmenis, kur kiekvienam išskirtam diktoriui daroma prielaida, kad visos diktorius reikšmės bus vienodos visuose tyrimuose įrašuose, bet kiekvienas kanalo veiksnys bus skirtingas kiekvienam įrašui.

Pagrindinė jungtinių faktorių analizės idėja, pateikta P. Kenny (2006) – surasti du po-erdvius, kurie atvaizduoja diktorius ir kanalo nuokrypius. M. Senoussaoui ir kt. (2010) eksperimentai parodo, kad jungtinių faktorių analizė yra tik dalinai sėkminga, atskiriant diktorius ir kanalo skirtumus. Kanalo erdvė turi papildomos naudingos informacijos diktorių atskyrimui. Minėtame darbe

yra siūloma suminio nuokrypio bendra erdvė, kurioje gali būti modeliuojami tiek diktoriaus, tiek kanalo nuokrypiai.

Autoriai A. Kanagasundaram ir kt. (2011) išskiria, kad ieškant patenkinamo efektyvumo, diktoriaus atpažinimo uždaviniuose buvo gana aktyviai koncentruojamasi į jungtinių faktorių analizės bei atraminėmis vektorių mašinomis pagrįstus diktoriaus atpažinimo užduočių sprendimus. Tačiau dauguma tyrimų parodė, kad šių sistemų efektyvumas gana stipriai krenta, kai yra dirbama su trumpais (<10 sekundžių) šnekos pavyzdžiais.

G. Lv ir H. Zhao (2012), naudodami 100 diktorių pavyzdžių garsyną, pasiekė 93 proc. diktoriaus identifikavimo tikslumą, naudodami JFA.

1.3.5 i-vektoriai

Pastaraisiais metais diktoriaus atpažinimo uždaviniui pradėti naudoti i-vektoriai (Lei ir kt. 2014; Richardson, Reynolds ir Dehak 2015). i-vektorius dar vadinamas tarpiniu vektoriumi,

$$\mu = m + Tw, \quad (10)$$

kur m yra universalus fono modelio super-vektorius, T yra mažos dimensijos matrica, kuri pažymi sumažinto bendrojo nuokrypio erdvę, o w yra nepriklausomas atsitiktinis vektorius su normaliuoju pasiskirstymu.

Šis metodas gali būti skirstomas į tris nuoseklius etapus: pakankamos statistikos surinkimas, i-vektorių skaičiavimas ir tiesinės diskriminantinės analizės (TDA) (angl. *linear discriminant analysis*) taikymas. Pakankamos statistikos surinkimas apibrėžiamas kaip procesas, kur eilė požymių vektorių (pvz. melų skalės kepstriniai koeficientai) yra vaizduojami kaip Baum-Welch algoritmo parametrų statistinės reikšmės, atitinkančios Gauso mišinių modelius, bei kartu atsižvelgiama į universaliojo fono modelį. Šie parametrai yra dauginamačiai, todėl jie turi būti transformuojami į vieną mažos erdvės (tarp 100 ir 1000) požymių vektorių (i-vektorių), kuriame yra informacija apie diktorių ir kitus kintamuosius. Apskaičiavus i-vektorių ir atlikus TDA, gauname atpažinimo rezultatą – i-vektoriaus įverčių palyginimą su kitais šnekos pavyzdžiais.

A. Kanagasundaram ir kt. (2011) išskiria, kad naudojant trumpos šnekos pavyzdžius, i-vektoriai gali identifikuoti diktorius geriau, nei kiti analizuoti algoritmai.

D. Garcia-Romero ir kt. (2014) savo išvadose teigia, kad i-vektorius, naudojantis Gauso mišinių modelius, surinkus pakankamą statistiką, gauna prastesnius rezultatus, nei taikant gilų neuroninį tinklą. Tiriant rezultatus, kurie būna gaunami valdomose aplinkose, gilus neuroninis tinklas rodo žymų našumo pagerėjimą, kuris detaliau yra apžvelgiamas 2-ame skyriuje.

1.3.6 Atraminių vektorių mašinos

Pasak J. Kamarauskos (2009), pagrindinė atraminių vektorių mašinų (AVM) (angl. *support vector machines*) idėja – dviejų skirtingų klasių mokymo duomenų projektavimas į aukštesnės dimensijos erdvę ir hiper-plokštumos, atskiriančios šias dvi klases aukštesnės dimensijos erdvėje. Hiper-plokštuma konstruojama maksimizuojant atstumą tarp dviejų artimiausių mokymo duomenų, priklausančių dviem skirtingoms klasėms.

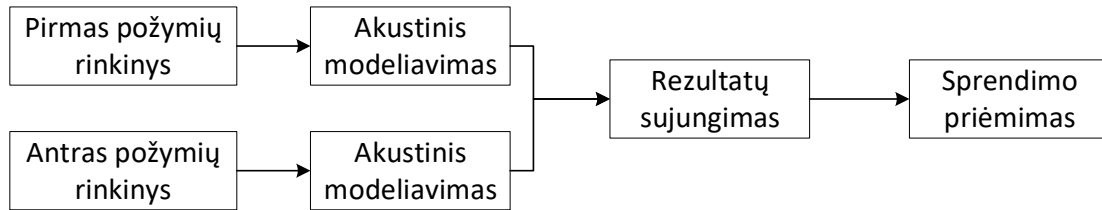
Matematiškai AVM uždavinio sprendimas yra aprašomas kaip kvadratinio optimizavimo problema (Cortes ir Vapnik 1995). AVM naudojami mokymo ir atpažinimo požymių vektoriai turi būti fiksuoto ilgio. J. Kamarauskas (2009) išskiria, kad V. Vapnik pasiūlytas atraminių vektorių mašinų klasifikavimo metodas leidžia pasiekti labai gerus rezultatus vaizdų apdorojime, ypač veido atpažinime. Eksperimentai rodo, kad šie klasifikatoriai veikia ne prasčiau už kitus, tuo tarpu reikalauja mažiau mokymo duomenų.

AVM panaudojimas taikytas diktoriaus atpažinimo uždaviniuose, atliktas šių autorių darbuose – P. J. Moreno ir P. P. Ho (2003), S. H. Chen ir Y. R. Luo (2009), G. Nijhawan ir M. K. Soni (2014) S. Mohanty ir B. K. Swain (2014), I. Ding, C. T. Yen ir D. C. Ou (2014). Minėtuose tyrimuose gauti rezultatai apibendrinami 3 lentelėje, skyriaus pabaigoje.

1.3.7 Požymių sintezės metodai

Siekiant užtikrinti atsparesnį ir patikimesnį klasifikavimo sprendimų priėmimą, gali būti taikomi požymių sintezės (PS) (angl. *feature fusion*) metodai

(Hasan, Jaafar ir Ramli 2012), pavyzdžiui, sujungiant MSKK požymius su subglotaliniais kepstriniais koeficientais (angl. *sub-glottal cepstral coefficients*) (Arsikere, Gupta ir Alwan 2014). Tokiu atveju požymių išskyrimas bus toks, kaip pavaizduota 11-ame paveiksle.



11 pav. Požymių sujungimo seka
Arsikere, Gupta ir Alwan 2014

Iš aukščiau esančios diagramos matyti, jog kiekvienas požymių rinkinys yra skaičiuojamas nepriklausomai nuo kito. Taip pat kiekvienam požymių rinkiniui atskirai yra atliekamas ir akustinis modeliavimas. Tačiau priimant galutinį sprendimą, rezultatai yra sujungiami.

Paprasčiausias požymių sujungimo variantas (Kinnunen ir Li 2009) užrašomas kaip:

$$s = \sum_{n=1}^{N_c} w_n s_n, \quad (11)$$

kur N_c yra klasifikatorių kiekis, w_n – n -tojo klasifikatoriaus svoris, s_n – n -tojo klasifikatoriaus rezultatas. Gautas rezultatas s toliau perduodamas klasifikatoriui priimti galutinį sprendimą.

Autoriai H. Arsikere, H. A. Gupta ir A. Alwan (2014) savo darbe nenurodo konkrečių naudotų požymių sujungimo metodų, tačiau sprendimų priėmimo etapuose sujungimui galima naudoti Bajeso teoremą, Dempster-Shafer išvadų generavimo metodus (Castanedo 2013).

1.3.8 Klasikinių diktoriaus identifikavimo metodų tikslumas

Šiame skyriuje apžvelgtų diktoriaus identifikavimo rezultatų apibendrinimas pateikiamas 2-oje lentelėje. Lentelėje pristatyti metodai, kurie buvo pritaikyti atliekant diktoriaus identifikavimo tyrimus. Kiti skyriuje apžvelgti, bet į lentelę neįtraukti klasifikavimo metodai – JFA, i-vektoriai, PS.

Kai kurie metodai neįtraukti dėl nepakankamo darbų, susijusių su diktoriaus identifikavimu, kiekio.

2 lentelė. PMM, GMM, UFM klasifikavimo metodų tikslumo palyginimas diktoriaus identifikavimo uždaviniui

Klasifikavimo metodas	Diktorių skaičius	Pasiektas identifikavimo tikslumas
PMM (vidurkis 97,80 proc.)	5	97,40 proc.
	20	84,50 proc.
	40	100,00 proc.
	630	98,09 proc.
GMM (vidurkis 87,30 proc.)	33	92,00 proc.
	44	99,25 proc.
	49	94,50 proc.
	100	nuo 77,00 proc.
UFM (vidurkis 74,12 proc.)	11	96,69 proc.
	22	95,45 proc.
	42	61,90 proc.
	50	67,50 proc.
	463	74,40 proc.
AVM (vidurkis 83,53 proc.)	8	95,00 proc.
	10	65,00 proc.
	26	67,10 proc.
	40	95,10 proc.
	50	84,70 proc.

Iš pateiktos lentelės matyti, kad PMM atpažinimo vidurkis yra aukščiausias iš palygintų metodų.

1.4 Diktoriaus identifikavimas taikant neuroninius tinklus

Kaip buvo rašoma 1.2 skyriuje, diktoriaus identifikavimo uždavinys susideda iš dviejų etapų: klasifikatoriaus apmokymo ir diktoriaus atpažinimo. Apmokymo etape iš balso įrašo išskiriami požymiai, kuriuos naudojant yra

apmokomas klasifikatorius. Atpažinimo etape iš balso įrašo išskyrus požymius, gauti požymiai lyginami su diktorių klasių požymiais ir priimamas identifikavimo sprendimas.

Neuroninių tinklų naudojimas diktoriaus atpažinimo uždaviniui yra sudėtingesnis nei klasikinių diktoriaus identifikavimo metodų taikymas, neuroniniam tinklui reikia daugiau mokymosi pavyzdžių (Lei ir kt. 2014). Kiti tyrimai giliųjų neuroninių tinklų srityje, kaip teigia Y. Lei ir kt. (2014), parodo, kad neuroninius tinklus galima pritaikyti ir akustiniam modeliavimui. Juose parodomas ryškus rezultatų pagerėjimas, neuroninius tinklus lyginant su Gauso mišinių modeliais.

Įvairių modelių atpažinimo (angl. *pattern recognition*) uždaviniuose, naudojant neuroninius tinklus, per pastaruosius 5 metus buvo gautas pastebimas atpažinimo tikslumo pagerėjimas, palyginus su tikslumu, gaunamu naudojant kitus metodus. Kaip nurodoma (Schmidhuber 2014), kalbos apdorojimo uždaviniuose, konkrečiai – fonemų klasifikavimo konkurse su TIMIT duomenų baze 2013-aisiais metais, geriausią tikslumą parodė neuroninis tinklas, nors prieš tai vykdytuose konkursuose visada geriausią tikslumą pasiekdavo paslėptais Markovo modeliais pagrįsti sprendimai.

Reikia atkreipti dėmesį, kad prieš tai neuroninių tinklų plitimas ir pritaikymas praktikoje buvo nepopuliarus dėl didelių skaičiavimo resursų poreikio tinklo apmokymui bei mokymosi pavyzdžių skaičiaus, kuris yra taip pat reikalingas efektyviam tinklo apmokymui.

1.4.1 Neuroniniai tinklai

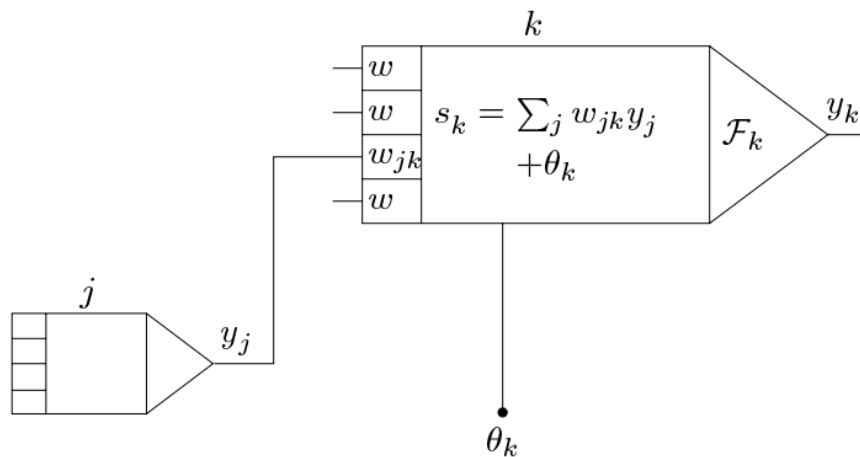
Diktoriaus akustiniam modeliavimui naudojamus klasifikatorius galima skirstyti į klasikinius, kurie detaliau buvo apžvelgti 1.2.3 skyriuje, ir neuroniniais tinklais paremtus klasifikatorius. Apžvelgti klasifikatoriai (išskyrus PS metodą) naudojo tikimybės tankio funkcijas, panaudojant mokymo pavyzdžius tikimybių tankio skaičiavimui. Mokant neuroninius tinklus, skirtingai – darome prielaidą, kad yra žinomos diskriminantinės funkcijos ir mokymo pavyzdžius naudojame šių funkcijų parametrų nustatymui.

Neuroniniai tinklai, kaip klasifikavimo metodas, naudojamas automatinėse kalbančiojo atpažinimo sistemose, išpopuliarėjo apie 1990 metus. Nuo tada neuroninius tinklus bandyta pradėti taikyti įvairiems kalbos atpažinimo uždaviniams spręsti, tokiems kaip fonemų klasifikavimui ir žodžių atpažinimui.

Neuroninių tinklų parametrai. Neuroniniai tinklai yra lygiagrečios paskirstytos funkcijos. Neuroninio tinklo modelyje išskirti pagrindiniai susiję aspektai yra (Krose ir Smagt 1996):

- Neuronas arba celė, j, k
- Aktyvacijos būseną y_j kiekvienam neuronui, kuri atitinka neurono išėjimą
- Jungtys tarp neuronų, nusakomos w_{jk} svorio koeficiento, kuris nustato j neurono poveikį k jungčiai
- Sklidimo taisyklė, nustatanti s_k
- Aktyvavimo funkcija F_k
- Įėjimo nuokrypis kiekvienam neuronų sluoksniui θ_k

Pagrindiniai paminėti neuroninio tinklo komponentai pavaizduoti 12-ame paveiksle.



12 pav. Pagrindinių neuroninio tinklo komponentų schema

B. Krose ir P. Smagt (1996)

Schemoje kairėje pusėje matomas įėjimo sluoksnis, kuris patenka į j neuroną. Pritaikius aktyvavimo būseną y_j , j -asis neuronas gauna svorio koeficientą w_{jk} . Iš to k -tajam neuronui apskaičiuojama sklidimo taisyklė s_k . Gautas rezultatas yra perduodamas aktyvavimo funkcijai $\mathcal{F}_k$, po kurios gaunama esamoji

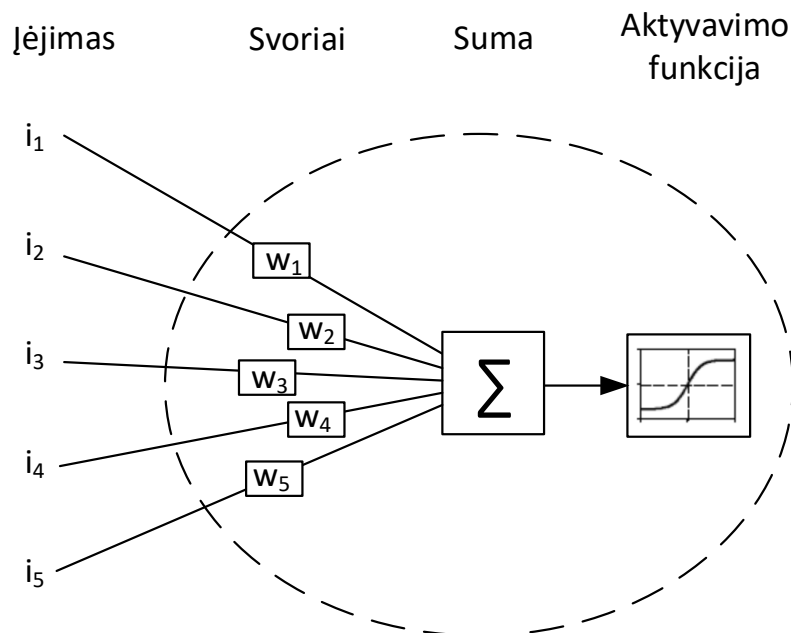
aktyvacija $y_k(t)$. Aktyvacija vykdoma sinchroniškai visam tinklui arba asinchroniškai laiku t .

Dažniausiai k -tojo neurono funkcija yra pasvertoji įėjimų suma su nuokrypiu θ , kuri yra užrašoma kaip:

$$s_k(t) = \sum_j w_{jk}(t) + \theta_k(t), \quad (12)$$

kur w_{jk} yra j -tojo neurono svorio koeficientas, θ – nuokrypis (angl. *bias*).

Perceptrono diagrama apibendrina patį elementariausią atvejį (13-ame paveiksle). Pavaizduotame tinkle yra 5 įėjimai $i_1, \dots, i_5$, kuriems yra pritaikomi svoriai $w_1, \dots, w_5$, sklaidimo taisyklė, atliekanti sumavimą, ir aktyvavimo funkcija.



13 pav. Neuroninio tinklo diagrama
Smith (2003)

Neuroninio tinklo hiper-parametrai. Neuroninio tinklo hiper-parametrai nurodo įvairius parametrus, kurie įtakoja atpažinimo rezultatus, tinklo mokymo spartą ir kt. Pagrindiniai neuroninių tinklų hiper-parametrai: tinklo topologijos parinkimas, neuronų kiekis sluoksnyje, tinklo sluoksnių kiekis, tinklo klaidos funkcijos optimizavimas ir mokymosi žingsnis (angl. *learning rate*).

Įėjimų skaičius yra nusakomas pagal įėjimo vektorių sudarančių požymių skaičių, o išėjimo sluoksnis – pagal klasių skaičių. Tuo tarpu paslėptų neuronų

skaičius nulemia klasifikavimo funkcijos kompleksiskumą ir jos sprendinių ribas. Jeigu klasės gerai atskiriamos arba gali būti atskiriamos tiesiškai, užtenka nedidelio neuronų kiekio. Ir priešingai, jei klasės yra persidengiančios, reikalingas didesnis neuronų kiekis (Duda, Hart ir Stork 2000).

1.4.1.1 Aktyvavimo funkcijos

Neuroninio tinklo aktyvavimo funkcija, kaip nurodo R. O. Duda, P. E. Hart ir D. G. Stork (2000) – tai įėjimo vektoriaus ir neurono išėjimo tarpusavio sąryšis. Kiekvienas paslėptas tinklo neuronas panaudoja įėjimo svorių sumą, kad aktyvuotųsi arba ne.

Tinklo aktyvavimo funkcijos dažniausiai yra monotoninės funkcijos, nustatančios tam tikrą slenkstį (Krose ir Smagt 1996):

$$\begin{aligned} y_k(t+1) &= \mathcal{F}_k(s_k(t)) = \\ &= \mathcal{F}_k\left(\sum_j w_{jk}(t)y_j(t) + \theta_k(t)\right), \end{aligned} \quad (13)$$

kur $\mathcal{F}_k$ – aktyvacijos funkcija, s_k – sklaidimo taisyklė, w_{jk} – j -tojo neurono svorio koeficientas, θ – nuokrypis.

Pagrindinės aktyvacijos funkcijos:

- Tiesinė:

$$\mathcal{F}(s_k) = s_k; \quad (14)$$

- Žingsninė:

$$\mathcal{F}(s_k) = \begin{cases} -1, & s_k < 0 \\ 1, & s_k \geq 0 \end{cases}; \quad (15)$$

- Sigmoidinė:

$$\mathcal{F}(s_k) = \frac{1}{1+e^{-s_k}}; \quad (16)$$

- Hiperbolinio tangento:

$$\mathcal{F}(s_k) = \frac{1-e^{-2s_k}}{1+e^{-2s_k}}; \quad (17)$$

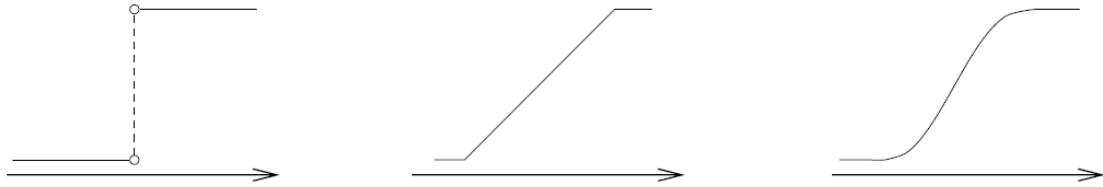
- Išlyginta tiesinė (ReLU):

$$\mathcal{F}(s_k) = \max(s_k, 0); \quad (18)$$

- Softmax:

$$\mathcal{F}(s_k)_j = \frac{e^{s_{kj}}}{\sum_{n=1}^N e^{s_{kn}}} ; \quad (19)$$

Žingsninės, tiesinės ir sigmoidinės funkcijos pavaizduotos žemiau esančiame grafike (14 paveikslas).



14 pav. Žingsninės, tiesinės ir sigmoidinės funkcijos diagramos

B. Krose ir P. Smagt (1996)

1.4.1.2 Neuroninio tinklo mokymas

Neuroninių tinklų mokymas yra tinklo neuronų jungčių svorių w_{jk} nustatymas pagal turimus mokymo pavyzdžius. Tikimasi, kad tinklas po mokymo, gavęs įėjimą, priimtų teisingą klasifikavimo sprendimą, t. y. aktyvuotų teisingą išėjimo sluoksnio neuroną. Kitaip tariant, kiekvienam įėjimo vektoriui x tikimės gauti rezultatą d , kuris būtų kuo arčiau tikrosios klasės y . Iš čia gauname delta taisyklę ($d - y$), kuri yra naudojama skaičiuojant nuostolio arba klaidos funkcijas, kad apskaičiuotume svorius (Krose ir Smagt 1996). Taigi, neurono svorio pokyčio skaičiavimas atliekamas pagal formulę:

$$\Delta w_{jk} = \gamma \delta_k y_j , \quad (20)$$

kur δ yra klaidos sklidimo signalas k neuronui, gaunamiam įėjimą ir išėjimą iš neurono j , γ yra mokymosi žingsnis.

Klaidos sklidimo signalas δ_o – indeksu pažymėto išėjimo sluoksnio neuronui yra užrašomas kaip:

$$\delta_o = (d_o - y_o) \mathcal{F}'(s_o) , \quad (21)$$

kur $\mathcal{F}'$ – aktyvacijos funkcijos išvestinė, s_o – sklidimo taisyklė.

Klaidos sklidimo signalas δ_h – indeksu pažymėto paslėpto sluoksnio neuronui yra užrašomas kaip:

$$\delta_h = \mathcal{F}'(s_h) \sum_{o=1}^{N_o} \delta_o w_{ho} , \quad (22)$$

kur $\mathcal{F}'$ – aktyvacijos funkcijos išvestinė, s_h – sklidimo taisyklė paslėptam sluoksniui, w_{ho} – svoriai paslėptam sluoksniui, N_o – paslėptų neuronų skaičius, h – paslėpto sluoksnio neuronas, o – išėjimo sluoksnio neuronas, δ_o – klaidos sklidimo signalas.

Kadangi pagrindinė aprašomo mokymo idėja yra pirmiausia apmokyti išėjimo sluoksnio neuronus, o paskui klaidą perduoti paslėptiems sluoksniams, šis mokymo būdas yra vadinamas klaidos sklidimo atgal (angl. *back-propagation*) mokymu. Mokymo tikslas – minimizuoti vidutinę kvadratinę klaidą tarp stebimo neuroninio tinklo išėjimo ir norimo išėjimo. Optimizavimui dažnai naudojamos atsitiktinio gradientinio nusileidimo tikslo funkcijos arba jų variantai, dėl greito gebėjimo mokytis naudojant didelį kiekį mokymo pavyzdžių (Richardson, Reynolds ir Dehak 2015; Keskar ir kt. 2017).

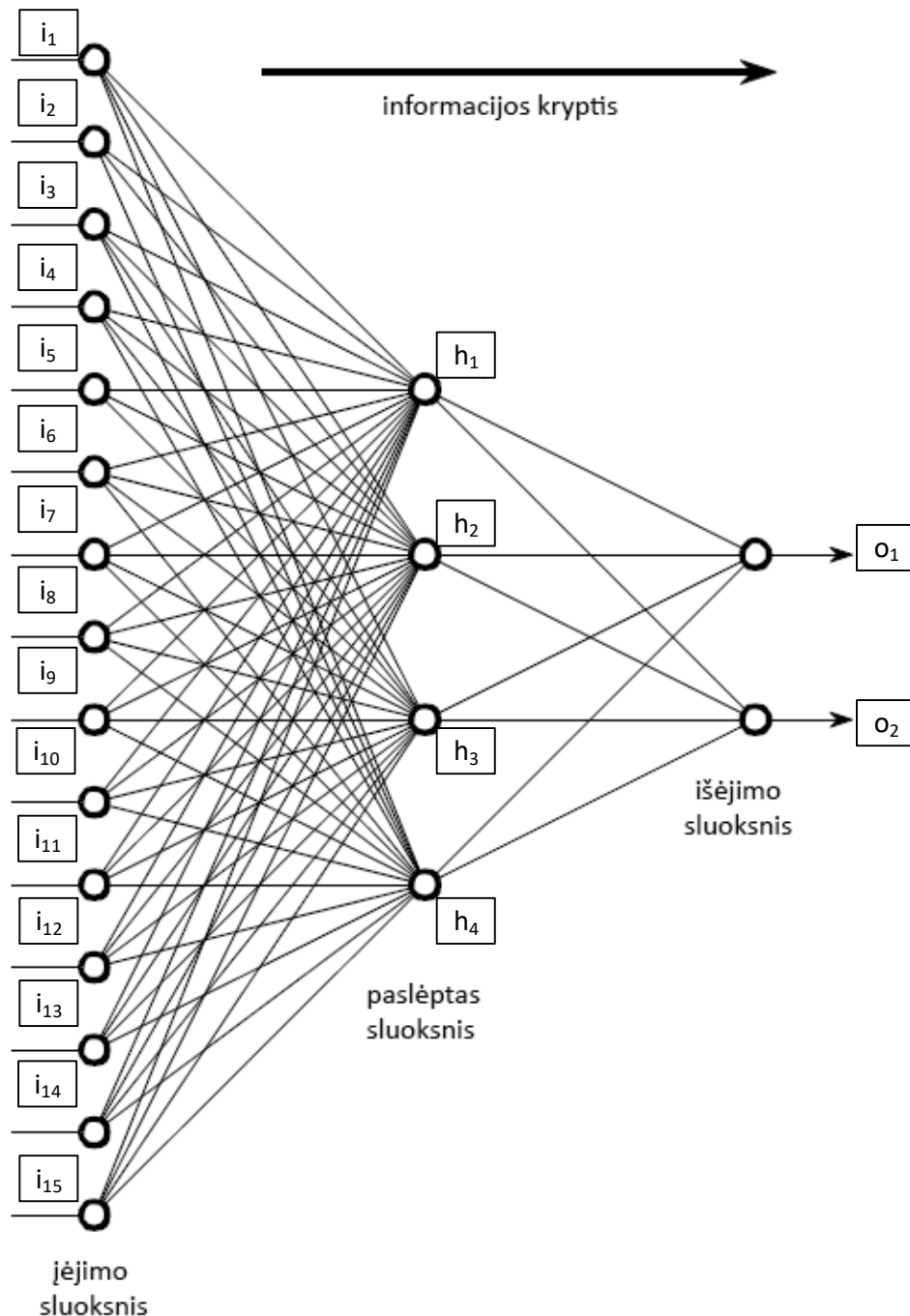
Mokymo tipai. Neuroninių tinklų apmokymas, kaip išskiria R. O. Duda, P. E. Hart ir D. G. Stork (2000), taip pat priklauso nuo to, koku režimu vykdome mokymą:

- Realus laiko (angl. *on-line*), kai tinklo svoriai yra pritaikomi kiekvienam mokymo pavyzdžiui
- Paketais (angl. *batch*), kai tinklo svoriai yra pritaikomi kiekvienam paketui, sudarytam iš mokymo rinkinio
- Tikimybinis, kai naudojami statistiškai reprezentatyvūs mokymo pavyzdžių rinkiniai

Kiekvienas iš paminėtų būdų turi savo privalumų ir trūkumų ir priklauso nuo pritaikymo būdo (Duda, Hart ir Stork 2000). Pavyzdžiui, realaus laiko mokymas vykdomas, kai turime mažai atminties, apdorojimas dalimis vyksta lėčiau, bet galima išnaudoti grafinių procesorių atminties ir paralelinio apdorojimo galimybes. Tipiškai apmokymas atliekamas duomenis imant paketais, kuriuose yra tarp 32 ir 512 pavyzdžių (Keskar ir kt. 2017).

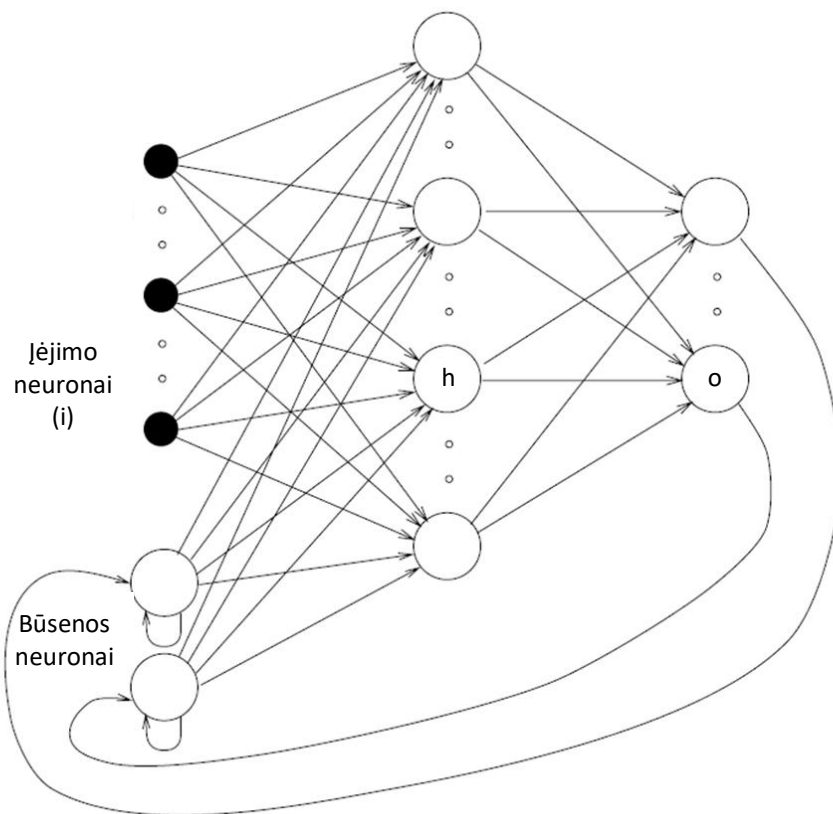
1.4.2 Neuroninio tinklo topologijos

Pagal neuronų sujungimo tipus gali būti išskirtos dvi pagrindinės topologijos (Krose ir Smagt 1996; Duda, Hart ir Stork 2000): perdavimo pirmyn tinklai (15-as paveikslas); rekurentiniai neuroniniai tinklai (16-as paveikslas).



15 pav. Tipinė dviejų sluoksnių, perdavimo pirmyn tipo, neuroninio tinklo diagrama

Smith (2003)



16 pav. Rekurentinis neuroninis tinklas su būsenos neuronais įėjimo sluoksnyje

B. Krose ir P. Smagt (1996)

Šių topologijų skirtumai:

- Perdavimo pirmyn tinklai, kur šiuo atveju duomenys tinklo viduje yra perduodami tik tolimesnėms aktyvacijoms ir nėra jokio grįžtamojo ryšio. Šio tipo neuroninio tinklo diagrama pateikiama 15-ame paveiksle.
- Rekurentiniai neuroniniai tinklai, kuriuose yra jungtys iš vėlesniuose sluoksniuose esančių neuronų į prieš tai einančių sluoksnių neuronus. Taip yra sukuriamas grįžtamasis ryšys tarp neuronų. Šio tipo neuroninio tinklo diagrama pateikiama 16-ame paveiksle, kur priešingai nei 15-ame paveiksle, matome, kaip iš išėjimo tinklo sluoksnio informacija yra perduodama į prieš tai ėjusio sluoksnio būsenos neuronus.

Kiekvienos iš šių topologijų tipai gali būti skirstomi į subtopologijas, priklausančias tam tikram neuroninio tinklo tipui.

Tinklo sluoksnių skaičius. Perdavimo pirmyn tipo tinkluose įėjimo sluoksnyje nėra atliekamų operacijų, todėl įėjimo sluoksnis į bendrą tinklo

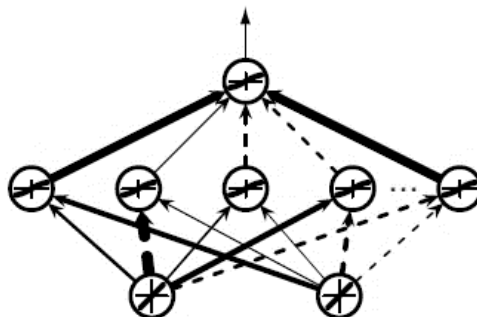
sluoksnių skaičių nėra įtraukiamas (Krose ir Smagt 1996). Taigi, tinklas, turintis vieną įėjimo sluoksnį (i), vieną paslėptą sluoksnį (h) ir vieną išėjimo sluoksnį (o), yra vadinamas dviejų sluoksnių tinklu ir yra atvaizduotas 15-ame paveiksle.

Gilūs neuroniniai tinklai. Bendro sutarimo, kas gali būti laikoma giliu neuroniniu tinklu, nėra, tačiau autoriai (Schmidhuber 2014; Deng ir Yu 2014), mini, kad giliu neuroniniu tinklu yra laikomas toks dirbtinis neuroninis tinklas, kuris turi daugiau negu du paslėptus sluoksnius. Minėti autoriai (Schmidhuber 2014; Deng ir Yu 2014) neįvardina, kiek sluoksnių turi būti giliame tinkle. Kaip pažymi Y. Lei ir kt. (2014), tipinis gilus neuroninis tinklas turi tūkstančius elementų kiekviename paslėptame sluoksnyje ir bent nuo 5 iki 7 paslėptų neuronų sluoksnių.

Gilaus neuroninio tinklo klasifikatorius dažniausiai yra daugiasluoksnis perceptronas, kuris detaliau aprašomas 1.4.2.1 skyriuje, tačiau, kaip pažymi J. Schmidhuber (2014), patys giliausi neuroniniai tinklai yra rekurentiniai neuroniniai tinklai. Dar viena papildoma savybė yra ta, kad gilaus neuroninio tinklo topologijoms reikalingos labai didelės mokymo pavyzdžių apimtys.

1.4.2.1 Daugiasluoksnis perceptronas

Didinant tinklo sudėtingumą ir tarp įėjimo ir išėjimo sluoksnių pridedant papildomą, paslėptą sluoksnį, gauname dviejų sluoksnių tinklą. Daugiau negu vieną sluoksnį turintys tinklai vadinami daugiasluoksniais perceptronais (angl. *multi-layer perceptron*) (DP). Dviejų sluoksnių tinklo pavyzdys pateikiamas 17-ame paveiksle.



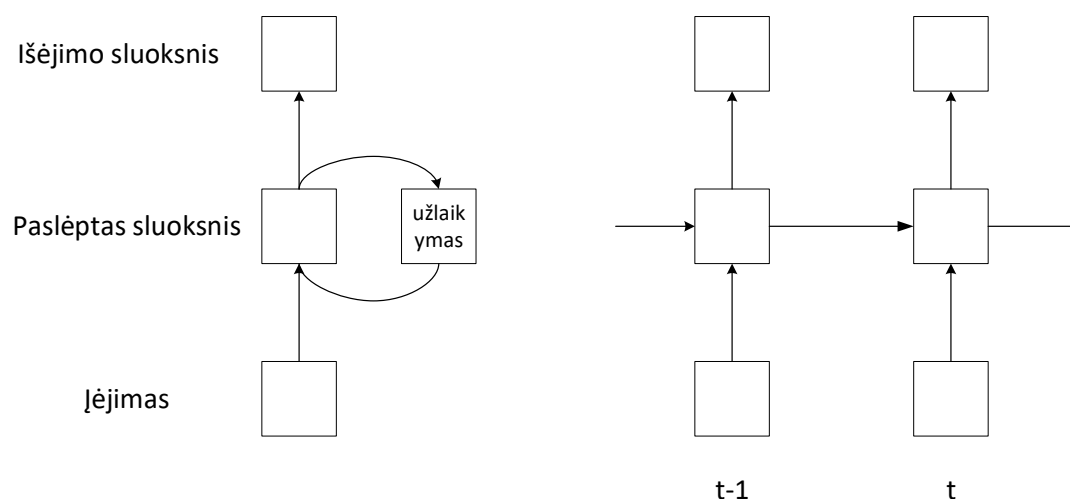
17 pav. Daugiasluoksnis perceptronas
Duda, Hart ir Stork (2000)

17-ame paveiksle pateikiamas neuroninio tinklo pavyzdys turi du įėjimo neuronus, paslėptas sluoksnis yra sudarytas iš 5 neuronų, o išėjimo sluoksnyje yra vienas neuronas. B. Krose ir P. Smagt (1996) teigimu, galima įrodyti, jog naudojant šią topologiją, pritaikius atitinkamus neurono svorius ir jungtis tarp neuronų, galima modeliuoti bet kokią tiesinį klasifikatorių.

Autoriai (Islam, Khan ir Haque 2013; Nath ir Kalita 2014; Chauhan 2017; Ge ir kt. 2017) savo darbuose pritaiko daugiasluoksnius perceptronus diktoriaus akustiniam modeliavimui. Apibendrinti minėtų autorių rezultatai pateikiami 3 lentelėje, skyriaus pabaigoje.

1.4.2.2 Rekurentiniai neuroniniai tinklai

Rekurentinis neuroninis tinklas (RNT) (angl. *recurrent neural network*) yra vienas iš neuroninių tinklų tipų, kurių dėka pastaruoju metu yra gaunami geri klasifikavimo rezultatai vaizdų ir šnekos atpažinimo srityse. Tokio neuroninio tinklo struktūra pateikiama 18-ame paveiksle. Jame parodyta, kaip informacija iš įėjimo sluoksnių yra perduodama į paslėptuosius tinklo sluoksnius ir kaip informacija sklinda laiko žingsniu t . Paslėptame tinklo sluoksnyje informacija taip pat yra perduodama toliau einantiems paslėpto tinklo neuronams (Schuster ir Paliwal 1997).



18 pav. Rekurentinio neuroninio tinklo struktūra ir informacijos sklaidimas tinkle laiko žingsniu t
Schuster ir Paliwal (1997)

Turint įėjimo vektorių seką $x=(x_1,...,x_T)$, RNT paslėpto sluoksnio vektoriaus $h = (h_1,...,h_T)$ ir išėjimo vektoriaus $y=(y_1,...,y_T)$ sekos randamos kartojant skaičiavimus:

$$h_t = \mathcal{F}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) , \quad (23)$$

$$y_t = W_{hy}h_t + b_y , \quad (24)$$

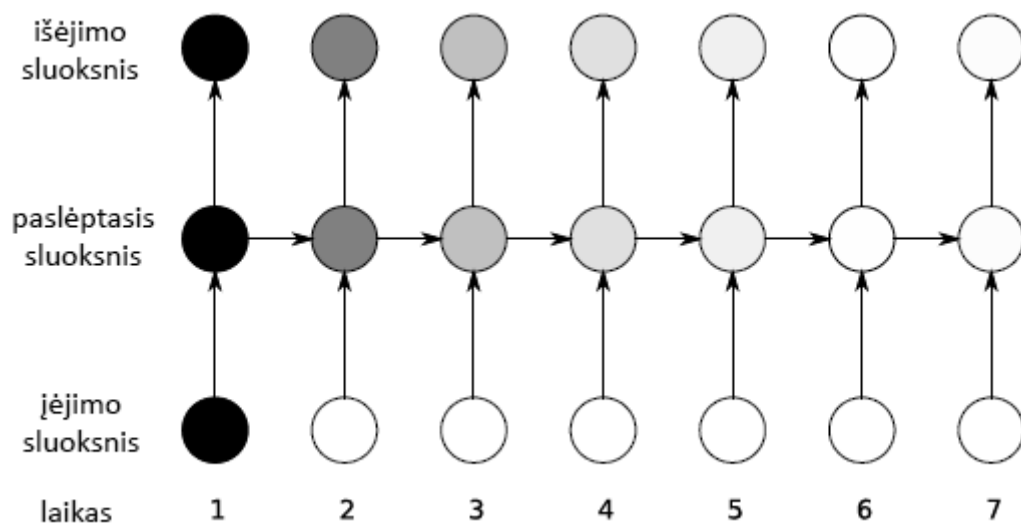
kur W indeksai nurodo svorių matricas, o b – nuokrypių vektorius ir $\mathcal{F}$ yra paslėpto sluoksnio aktyvacijos funkcija (dažniausiai naudojama sigmoidė, pagal (16) formulę) (Graves, Mohamed ir Hinton 2013).

RNT, priešingai nuo DP tipo tinklų, įėjimo sluoksniui gali naudoti kintančių dimensijų įėjimo vektorius (Schuster ir Paliwal 1997). Ši savybė suteikia geresnes sąlygas RNT panaudoti natūralios kalbos supratimui bei apdorojimui, tame tarpe kalbos generavimui, vaizdo apdorojimui ir įvairioms kitoms užduotims atlikti (Gal ir Ghahramani 2016).

Dažnai neuroninių tinklų neigiama savybė yra per didelis tinklo prisitaikymas prie mokymo duomenų. Rekurentiniai neuroniniai tinklai leidžia sukurti modelius, rodančius ypatingą našumą, tačiau tuo pačiu yra ir didelė per didelio prisitaikymo prie mokymo duomenų grėsmė. Generalizavimo trūkumas yra vienas iš pagrindinių trikdžių, kuris apriboja tokio tipo neuroninių tinklų naudojimą. Tai ypač svarbu tinkluose, kurių mokymui naudojama mažai apmokymo duomenų. Kaip teigia Y. Gal ir Z. Ghahramani (2016), per didelis prisitaikymas gali įvykti net ir naudojant neuronų atjungimo (angl. *dropout*) techniką, kuri detaliau paaiškinta 1.4.3 skyriuje.

Pasak A. Graves, A. Mohamed ir G. Hinton (2013), RT yra įkvėpta neuroninių jungčių cikliškumo ir naudoja iteratyvų neuronų apmokymo ciklą informacijos išsaugojimui. Todėl jie turi keletą naudingų savybių: šie tinklai gali lanksčiai apibendrinti kontekstinę informaciją ir dėl kontekstinės informacijos apibendrinimo galimybių RT mokymui galima naudoti skirtingų įėjimo duomenų rinkinius.

RT taip pat turi „užmaršumo“ savybę, kuri iliustruota 19 paveiksle, kaip per laiko momentus blankstanti neurono aktyvacija. Čia pavaizduota, kaip, einant laikui, neurono aktyvacijos paslėptuose sluoksniuose silpsta.

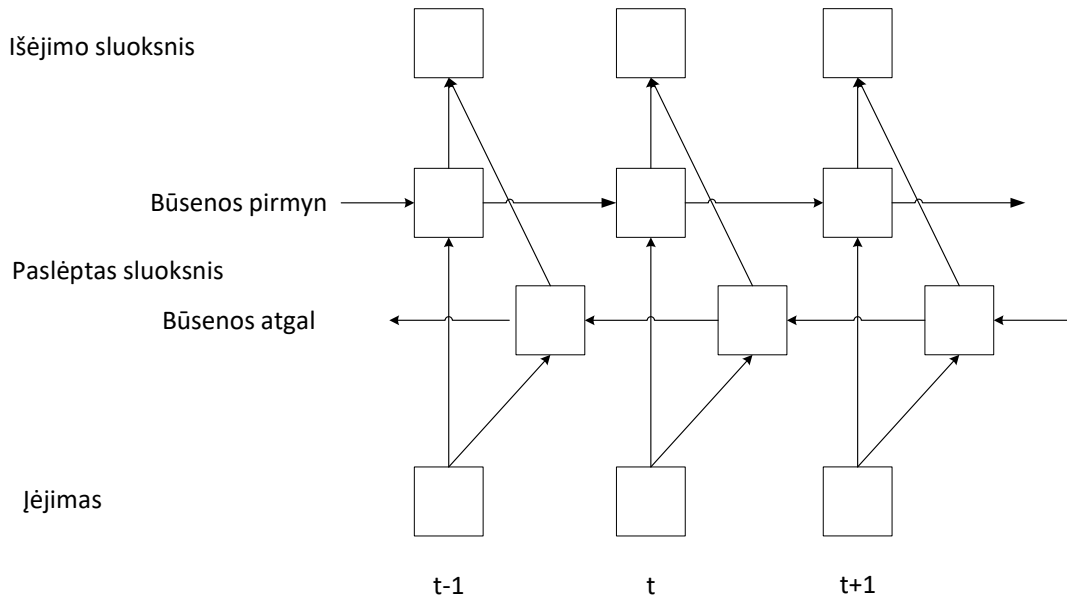


19 pav. Rekurentinių neuroninių tinklų užmaršumo savybė
Graves (2012)

Kaip pažymi A. Graves, A. Mohamed ir G. Hinton (2013), GRT tinklai buvo sukurti vienmatėms sekoms, tačiau kai kurios jų savybės, pavyzdžiui, kontekstinių savybių apibendrinimas ir konteksto panaudojimas yra labai reikalingos sprendžiant vaizdo ir garso įrašų klasifikavimo uždavinius.

Dvikryptis rekurentinis tinklas. Klasikinio RT patobulinimą pasiūlė M. Schuster ir K. K. Paliwal (Schuster ir Paliwal 1997), pavadindami jį dvikrypčiu RNT (DRNT). Pagrindinė autorių darbe (Schuster ir Paliwal 1997) pristatomo patobulinimo idėja yra tokia, jog RNT tinkle esančios sekos gali atsižvelgti ne tik į žingsnius į ateitį, bet ir į laiko žingsnius į praeitį. Bendra šio patobulinimo struktūra yra atvaizduojama 20-ame paveiksle.

20-ame paveiksle matome, kaip informacija iš įėjimo sluoksnio yra perduodama į paslėptą sluoksnį. Paslėptame sluoksnyje vienas neuronas informaciją perduoda prieš tai buvusiam laiko momentui, kitas – būsimam laiko momentui.



20 pav. Dvikrypčio rekurentinio neuroninio tinklo struktūra
Schuster ir Paliwal (1997)

Turint įėjimo vektorių seką $x=(x_1,...,x_T)$, skaičiuojant dvikryptį RNT, paslėpto sluoksnio išėjimo apskaičiavimas būsimu laiko momentu $\vec{h}_t$, prieš tai buvusiu laiko momentu $\vec{h}$ ir išėjimo vektorius y atliekamas kartojant skaičiavimus, nuo $t = T$ iki t buvusio laiko momento skaičiavimui ir $t = 1$ iki T būsimam laiko momentui ir tada atnaujinant išėjimo sluoksnį:

$$\vec{h}_t = \mathcal{F}(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}), \quad (25)$$

$$\vec{h}_t = \mathcal{F}(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t+1} + b_{\vec{h}}), \quad (26)$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\vec{h}y}\vec{h}_t + b_y, \quad (27)$$

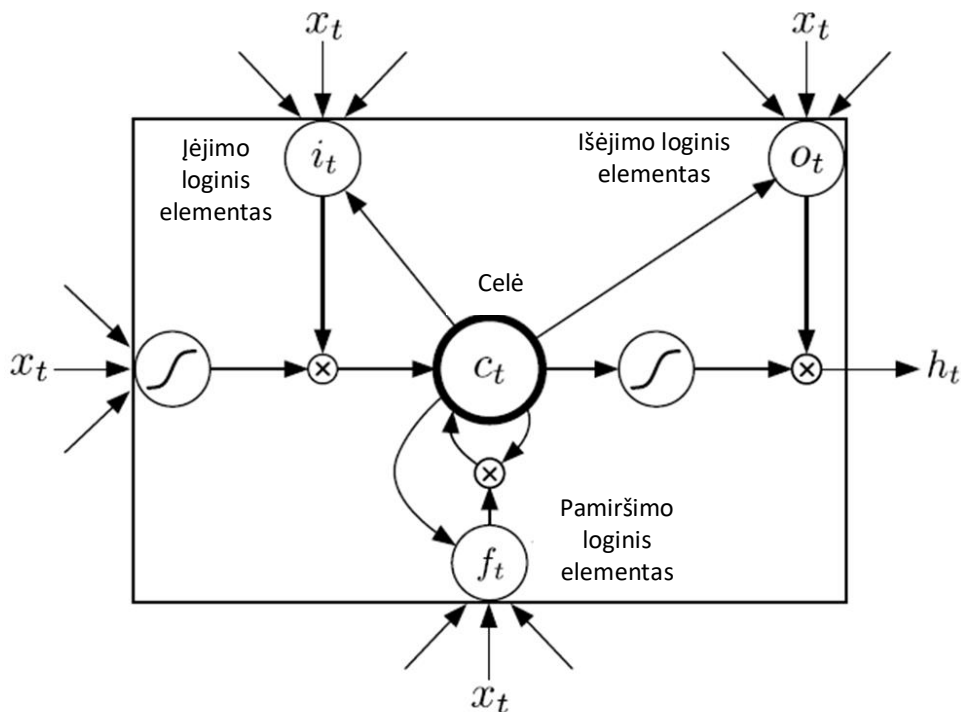
kur W nurodo svorių matricas, o b – nuokrypių vektorius ir $\mathcal{F}$ yra paslėpto sluoksnio aktyvacijos funkcija (Graves, Mohamed ir Hinton 2013).

Rekurentiniai tinklai naudojami diktoriaus identifikavimui G. Bhattacharya, J. Alam, T. Stafylakis ir P. Kenny (2016) darbe. Čia naudojamas vienasluoksnis RNT tinklas, turintis nuo 100 iki 700 neuronų: pasiekiamas iki 57 proc. diktoriaus identifikavimo tikslumas. Šiame darbe autoriai taip pat tiria ir 2 bei 3 sluoksnių RNT tinklų našumą, tačiau gaunami rezultatai yra prastesni nei 1 sluoksnio RNT tinklo. Palyginimui tie patys autoriai (Bhattacharya ir kt. 2016) atliko bandymus ir su dvikrypčiu RNT tinklu. Šiuo atveju pasiekti geresni rezultatai – 65 proc. diktoriaus identifikavimo tikslumas.

1.4.2.3 Ilgos trumpalaikės atminties tinklai

Ilgos trumpalaikės atminties (ITA) (angl. *long short-term memory*) tinklas – tai RNT tinklo tipas, 1997 metais pasiūlytas S. Hochreiter ir J. Schmidhuber (Hochreiter ir Schmidhuber 1997). Jo pagrindinės savybės pritaikytos taip, kad būtų išvengta RNT trūkumų, susijusių su svyruojančiais tinklo svoriais mokymo metu, bei kad būtų galimybė įvertinti ilgalaikę kontekstinę informaciją. Keletas iš pritaikymo pavyzdžių – kalbos apdorojimas, muzikos kompozicija ir kt. (Schmidhuber 2014).

ITA naudoja grįžtamąjį ryšį vadinamoje atminties celėje. Taip yra išsaugomi vėliausių įėjimų signalai, kurie tinkle sukuria „trumpalaikę atmintį“. Kiti atminties celės komponentai – įėjimo (angl. *input gate*), išėjimo (angl. *output gate*) ir užmiršimo loginiai elementai (angl. *forget gate*) – išsaugo šią trumpalaikę atmintį ilgesnį laiko tarpą, o tai padeda įvertinti ilgalaikę kontekstinę įėjimo informaciją. Tipinė ilgos trumpalaikės atminties tinklo celė yra pavaizduota 21 paveiksle.



21 pav. Ilgos trumpalaikės atminties tinklo celė
Graves, Mohamed ir Hinton (2013)

Pateiktoje diagramoje (21 pav.) parodyta, kaip įėjimo vektoriai x_t yra paduodami į ITA atminties celę ir yra naudojami kaip celės įėjimas. Loginių įėjimo i_t , išėjimo o_t ir pamiršimo f_t elementų įėjimai bei celė c_t yra vienodo ilgio paslėpto tinklo vektoriai. Išėjimas h_t – tai neuroninio tinklo paslėpto sluoksnio išėjimas. ITA komponentai yra aprašomi (Graves, Mohamed ir Hinton 2013) kaip:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}h_{t-1} + b_i), \quad (28)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}h_{t-1} + b_f), \quad (29)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (30)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}h_{t-1} + b_o), \quad (31)$$

$$h_t = o_t \tanh(c_t), \quad (32)$$

kur σ yra logistinė sigmoidinė funkcija ir i_t, o_t, f_t, c_t yra įėjimo, išėjimo, pamiršimo loginių elementų ir celės aktyvacijos vektoriai, $W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{xc}, W_{hc}, W_{cf}, W_{xo}, W_{ho}, W_{co}$ – svorių koeficientų matricos, b_i, b_f, b_c, b_o – nuokrypio vektoriai.

Dvikryptis ilgos trumpalaikės atminties tinklas. Dvikryptis ITA (DITA) tinklas yra skaičiuojamas pritaikius dvikrypčio RNT tinklo skaičiavimo principus ir idėjas ITA skaičiavimui tinklo (Graves, Mohamed ir Hinton 2013). Abiejų krypčių sujungimas atliekamas naudojant (27) formulę, atitinkamai apskaičiuojant (32) taikant praėjusį laiką ($t-1$) ir būsimą laiką ($t+1$).

A. Graves, A. Mohamed ir G. Hinton (2013) teigimu, uždaviniuose, kuriuose visą reikalingą įėjimo informaciją turime iškart, pavyzdžiui, balso ištaramui (angl. *utterance*), verta taip pat panaudoti būsimą laiko kontekstinę informaciją.

Ilgos trumpalaikės atminties tinklo panaudojimas su balso atpažinimu susijusiuose uždaviniuose. Fonemų atpažinimui naudotų ITA tinklo tikslumo rezultatai pateikti (Graves ir Schmidhuber 2005) darbe, kur išskirtas geriausias rezultatas naudojant ITA tinklą buvo 66 proc. teisingai identifikuotų fonemų, o naudojant DITA tipo tinklą geriausi rezultatai testavimo duomenims buvo 70,2 proc. teisingai identifikuotų fonemų. Minėti tyrimai buvo

atlikti su garsynu, kur ištarimus įrašė 184 unikalūs diktoriai, o duomenų bazė buvo padalinta į 3696 mokymo ir 1344 testavimo imtis. Šio tyrimo etape buvo išskirti MSKK požymiai, kurie toliau buvo naudojami atliekant neuroninio tinklo mokymą ir įvertinimą. ITA tinklo aktyvavimui naudota logistinė sigmoidinė funkcija.

ITA tinklas taip pat buvo taikytas identifikuojant diktoriaus kalbą. Autoriai (Zazo ir kt. 2016) pasiekė 70,9 proc. diktoriaus kalbos identifikavimo tikslumą, atliekant tyrimą su 40 skirtingų kalbų.

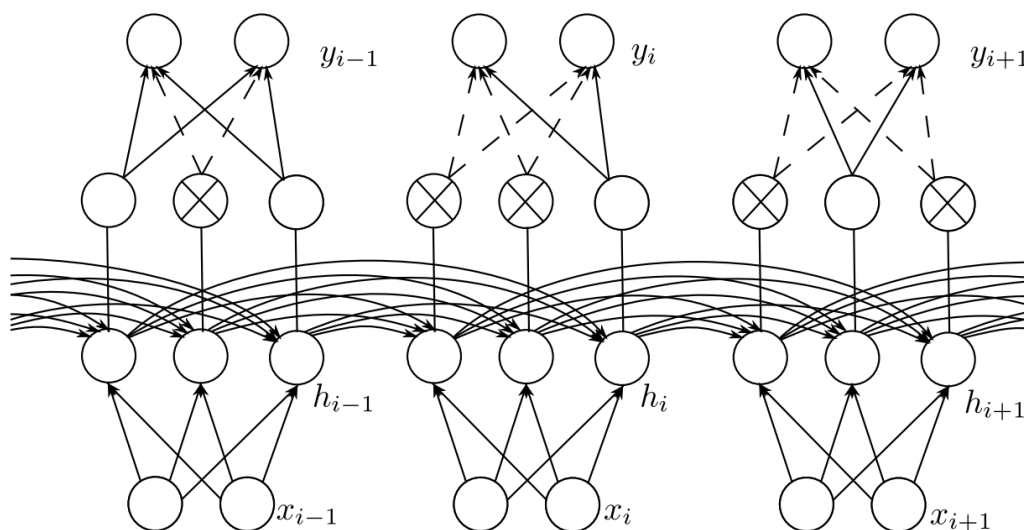
1.4.3 Apibendrinimo problema

Vienas iš sunkesnių su neuroniniais tinklais susijusių iššūkių – tinklo savybė, kad tinklas per daug prisitaikys prie mokymo duomenų. Siekiant išvengti per didelio prisitaikymo, reikia nuspręsti, kokio sudėtingumo tinklas turėtų būti naudojamas ir parinkti tinklo konfigūraciją.

Kitaip tariant, turint per daug neuronų, tinklo apibendrinimo savybės bus prastos, tačiau jei jų bus per mažai, tinklas per daug generalizuos (Duda, Hart ir Stork 2000). Vienas iš būdų giliame neuroniniame tinkle atlikti geresnį apibendrinimą (generalizuoti) yra atsitiktinis neuronų atjungimas.

Atsitiktinis neuronų atjungimas. Neuronų atjungimas, kaip išskiria Y. Gal ir Z. Ghahramani (2016), yra būdas, naudojamas siekiant generalizuoti neuroninį tinklą. Šiuo principu kiekvienos mokymo iteracijos metu atsitiktinio parinkimo būdu iš tinklo sluoksnio yra pašalinama dalis neuronų.

Kaip pažymi autoriai N. Srivastava ir kt. (2014), atsitiktinis neuronų atjungimas iš neuroninio tinklo gali padėti ir kai neturime didelio mokymo pavyzdžių kiekio. Pavyzdžiui, jei apmokymui skirti duomenys turi triukšmo, tinklas gali išmokti atpažinti būtent jį. Vykdam atpažinimą, tinklo tikslumas pastebimai kris. Rekurentinio neuroninio tinklo tinklo topologija su atjungimo neuronais pateikiama 22 paveiksle. Atsitiktinai atjungiami neuronai pažymėti X.



22 pav. Neuroninis tinklas su neuronų atjungimu

V. Pham ir kt. (2014)

1.4.4 Neuroninių tinklų diktoriaus identifikavimo metodų tikslumas

Nėra atlikta daug diktoriaus identifikavimo tyrimų naudojant neuroninius tinklus, tačiau kad suprastume galimą neuroninių tinklų potencialą diktoriaus identifikavimo uždavinyje, galime apibendrinti skirtingų autorių darbus šia tema (3-ia lentelė).

3 lentelė. DP, AVM neuroninių tinklų klasifikavimo tikslumo palyginimas diktoriaus identifikavimo uždaviniui

Neuroninio tinklo tipas	Diktorių skaičius	Pasiektas identifikavimo tikslumas
DP	6	Nuo 95,00 proc.
	8	87,50 proc.
	30	50,00 proc.
	83	91,00 proc.
	200	93,29 proc.

Apie kitų skyriuje pateiktų neuroninių tinklų tipų pritaikymą diktoriaus identifikavimo uždaviniui pakankamos informacijos rasti nepavyko. Tačiau neuroninių tinklų pritaikymą akustiniam modeliavimui, kai į neuroninio tinklo įėjimą yra paduodami MSKK požymiai, yra atlikę A. Graves ir J. Schmidhuber (2005). Šiame tyrime buvo atliekamas fonemų atpažinimas (4-ta lentelė). Iš

pateiktos lentelės matyti, jog didžiausias tikslumas buvo pasiektas naudojant ITA ir DITA neuroninių tinklų tipus.

4 lentelė. Fonemų atpažinimo tikslumo rezultatų palyginimas, naudojant neuroninio tinklo klasifikatorius. Graves ir Schmidhuber (2005)

Neuroninio tinklo topologija	Atpažinimo tikslumas	
	Mokymo imtis (3696 fonemos)	Testavimo imtis (1344 fonemos)
DP	67,60 proc.	63,10 proc.
RNT	69,90 proc.	64,50 proc.
DRNT	76,00 proc.	69,00 proc.
ITA	77,60 proc.	66,00 proc.
DITA	78,60 proc.	70,20 proc.

1.5 Skyriaus išvados

Šiame skyriuje apžvelgtas diktorius biometrinio atpažinimo uždavinys, ypatingą dėmesį skiriant diktorius identifikavimui pagal balsą. Pateikti tikslumo įvertinimo matai. Apžvelgtas diktorius identifikavimo procesas ir išskirti pagrindiniai žingsniai, reikalingi norint vykdyti diktorius atpažinimą. Taip pat pateikti kai kurie rezultatai, gauti taikant neuroninius tinklus panašaus tipo uždaviniui – šnekos atpažinimui.

Atlikus literatūros analizę, buvo gautos šios išvados:

- Pastebėta, jog dažniausiai yra naudojamas MSKK požymių rinkinys, kurį sudaro 39 požymių vektorius, sudarytas iš 13 MSKK, 13 delta ir 13 delta-delta požymių.
- Pastebima tendencija, kad diktorius identifikavimo tikslumas mažėja didėjant diktorių kiekiui. Tiesiogiai palyginti skirtingų autorių naudojamus klasifikavimo metodus yra sudėtinga dėl įrašo kokybės ir skirtingų balso įrašymo sąlygų. Taip pat tikslumo rezultatai priklauso nuo naudojamų balso signalo požymių rinkinio ir klasifikatorių.
- Neuroninių tinklų taikyme diktorius atpažinimui pastebėta, jog nėra išsamių diktorius identifikavimo tyrimų, kuriuose būtų palyginti

identifikavimo tikslumo parametrai, naudojant skirtingų tipo rekurentinių neuroninių tinklų topologijas.

- Apžvelgus mokslininkų fonemų identifikavimo tyrimus, galima teigti, jog didžiausias fonemų atpažinimo tikslumas gaunamas naudojant rekurentinio neuroninio tinklo topologijas. Tikėtina, kad diktoriaus identifikavimo uždaviniui gali būti stebimi panašūs tikslumo pagerėjimai, tačiau dėl mažo tyrimų kiekio šios išvados pagrindimui reikia papildomų diktoriaus identifikavimo eksperimentų.

2 DIKTORIAUS IDENTIFIKAVIMO METODAI

Šiame skyriuje yra aprašomi algoritmai ir metodai, siūlomi naudoti diktoriaus identifikavimui. Pirmoje skyriaus dalyje yra apžvelgiama modelių hiper-parametrų paieška. ITA tinklo hiper-parametrų paieškos kriterijai parinkti, atsižvelgiant į autorių (Greff ir kt. 2015) atliktus tyrimus akustinio modeliavimo kadro lygio klasifikavimo problemoms.

Antrame skyriuje yra detalizuojami diktoriaus identifikavimo klasifikatoriaus metodai: MSKK požymių išskyrimo algoritmas, PMM klasifikatoriaus mokymas, RNT klasifikatoriaus mokymas, rezultatų analizės ir palyginimo išskyrimas. Trečioje skyriaus dalyje yra pateikiamas apibendrintas siūlomas lietuviškai kalbančio diktoriaus identifikavimo metodas.

Tarpiniai skyriuje pateikiamų algoritmų ir hiper-parametrų paieškos rezultatai yra publikuoti moksliniame žurnale (Dovydaitis ir Rudžionis 2018).

2.1 Klasifikavimo metodų hiper-parametrų paieška

Iš visų atliktų tyrimų apžvalgos, pasirinkta atlikti tyrimus, paremtus dviem variantais: atliekant akustinį modeliavimą vienu iš efektyviausių kelis dešimtmečius naudojamų algoritmų – naudojant paslėptuosius Markovo modelius, bei atliekant akustinį modeliavimą su rekurentiniais neuroniniais tinklais, o tiksliau – šios architektūros modifikacija – ilgos trumpalaikės atminties tinklais. Kadangi iš kitų mokslininkų darbų buvo matyti, jog MSKK požymių išskyrimas yra populiariausias, dažniausiai sutinkamas bei gaunami rezultatai yra geri, priimtas sprendimas naudoti šį požymių rinkinį diktoriaus identifikavimo uždaviniui. Toks pasirinkimas taip pat leidžia įvertinti tik klasifikavimo metodo įtaką diktoriaus identifikavimo tikslumui.

Tiksliausiai atpažinimą vykdančio modelio parinkimui bus naudojamas tinklo paieškos (angl. *grid search*) metodas. PMM hiper-parametrų paieška bus atliekama naudojant kitų autorių darbuose dažniausiai minimus PMM būsenų skaičius – 5, 7, 10, 16, 22. Atlikus tyrimus bus nustatytas tiksliausiai diktorių identifikuojantis PMM būsenų variantas. Tuomet bus ištirti jo artimiausi kaimynai – -2, -1, 1, 2. Būsenų skaičius detalčiau aprašomas 1.3.1 skyriuje.

NT hiper-parametrų parinkimas. Su ITA tipo tinklais atliktas išsamus hiper-parametrų įtakos atpažinimo uždaviniams tyrimas aprašomas K. Greff ir kt. (2015) darbe. Tyrimo išvadose nurodoma, jog rezultatai tarp klasikinio (angl. „vanilla“) ITA tinklo topologijos ir tirtų 8 alternatyvų skiriasi nežymiai. Pagrindiniai K. Greff ir kt. (2015) tyrime nurodomi hiper-parametrų režiai, kurie apmokymo rezultatus leidžia suskaičiuoti sąlyginai greitai, nepakenkiant tikslumui, yra:

- Tinklo mokymo sparta: 0.01 – 0.0001
- Neuronų skaičius: 80 - 160

Apibendrinant aukščiau išvardintus kriterijus, sudaryta 5 lentelė. Lentelėje pateikiamos hiper-parametrų reikšmės, kiekvienam iš pasirinktų klasifikatorių. Pradinės reikšmės stulpelis nurodo parametrą, su kuriuo bus pradedamas vykdyti eksperimentas. Patikslinimo stulpelis nurodo, kaip bus pakeista didžiausią identifikavimo tikslumą pasiekusi pirminė hiper-parametro reikšmė.

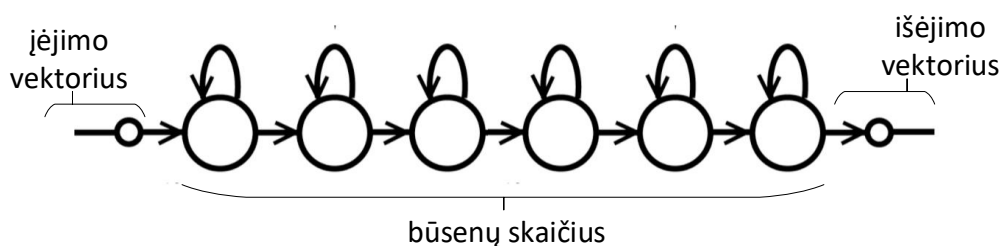
Kadangi PMM turi mažiau kintamų parametrų, jo efektyvumą galima įvertinti vienu etapu. Tuo tarpu RNT tinklo topologijos kūrimas priklauso nuo parametrų bei hiper-parametrų reikšmių, kur optimalių parametrų paiešką yra patogiau išskirstyti į keletą etapų.

5 lentelė. Klasifikatorių hiper-parametrų paieškos lentelė

Eksperimentas	Hiper-parametrų paieška	
	Pradinė reikšmė	Patikslinimas
PMM	Būsenų skaičius 5; 7; 10; 16; 22;	Prie tiksliausio -2;-1;+1;+2;
RNT 1 etapas	Tinklo topologija ITA 80;	Tinklo topologija DITA 80;
RNT 2 etapas	Neuronų atjungimo santykis: 0,2; 0,4; 0,6;	Prie tiksliausio -0,1; +0,1;

Eksperimentas	Hiper-parametrų paieška	
	Pradinė reikšmė	Patikslinimas
RNT 3 etapas	Neuronų skaičius didinamas 80 neuronų žingsniu, kol rezultatas nebegerėja	Gylis padidinamas +1 sluoksniu, kol rezultatas nebegerėja

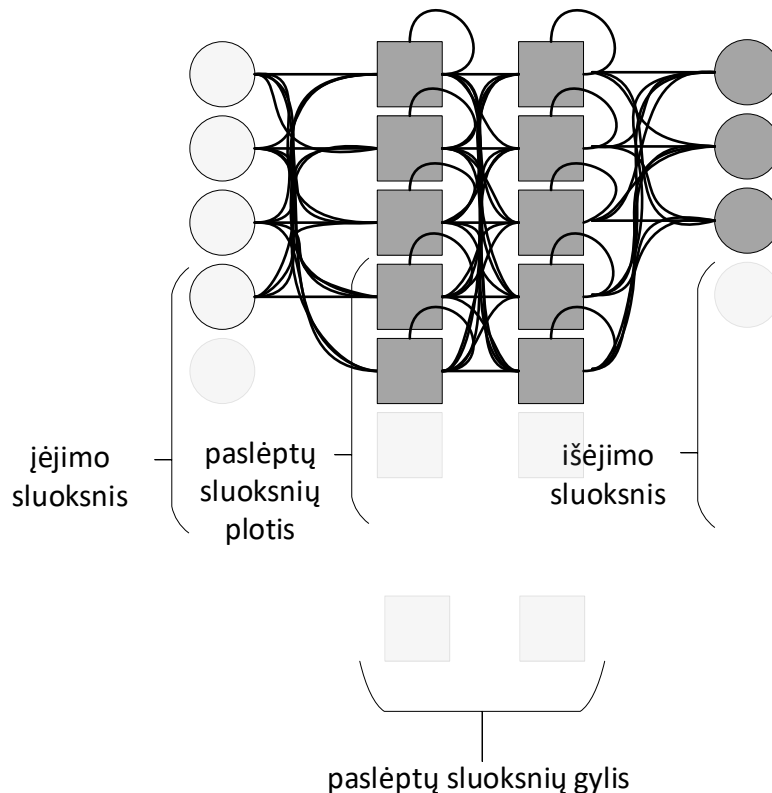
Principinė paslėpto Markovo modelio schema parodyta 23-ame paveiksle. Čia matoma bendrinė paslėpto Markovo modelio grandinė su 6 paslėptomis būsenomis. Ieškant tiksliausią identifikavimą atliekančio būsenų skaičiaus, ši grandinė bus trumpinama arba ilginama pagal 5-oje lentelėje nurodytus parametrus.



23 pav. Principinis paslėptas Markovo modelis

Principinė rekurentinio neuroninio tinklo topologija parodyta 24-ame paveiksle. Identifikavimo tikslumui padidinti sieksime rasti suboptimalius hiper-parametrus, išvardintus 5-oje lentelėje.

Pateikiamame paveiksle matome bendrinę RNT neuroninio tinklo topologiją su 4 įėjimo neuronais, 2 paslėptais sluoksniais, kur kiekvienas iš jų turi po 5 neuronus ir išėjimo sluoksnį su 3 neuronais. Atitinkamas įėjimo neuronų skaičius bus keičiamas pagal įėjimo vektoriaus matmenų kiekį, paslėpto sluoksnio neuronų plotis ir gylis – pagal 5-oje lentelėje nurodytus parametrus, o išėjimo sluoksnio neuronų kiekis – pagal klasifikuojamų klasių skaičių.



24 pav. Principinė naudojamo neuroninio tinklo topologija

Tyrimų metu rasti diktoriaus identifikavimo tikslumo parametrai yra nustatomi pagal (4) formulę.

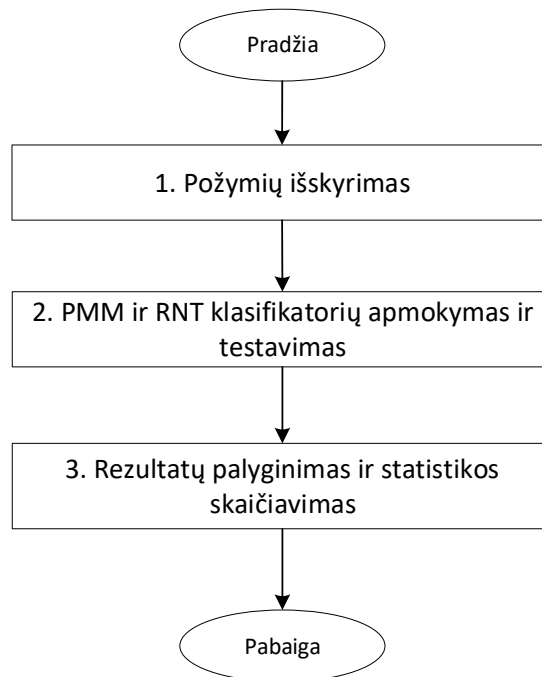
2.2 Identifikavimo tikslumo tyrimas

Identifikavimo tikslumo tyrimo veiksmų seka atvaizduota 25-ame paveiksle. Tyrimo veiksmų sekoje numatyti trys etapai. Pirmas – tai požymių išskyrimo etapas, kuris yra bendras visiems eksperimentams. Antras etapas – PMM ir RNT klasifikatorių apmokymas – yra padalintas į du sub-etapus, kuriuose atliekama tiksliausio klasifikatoriaus paieška:

1. PMM tyrime – ieškant tiksliausiai identifikuojančio diktorių PMM modelio, ieškomas būsenų skaičius, leidžiantis tiksliausiai identifikuoti diktorių.
2. RNT tyrime – parenkant skirtingas NT architektūras, ieškoma tinklo topologijos, kuri leistų tiksliausiai identifikuoti diktorių. Iš atliktos literatūros analizės galima numanyti, jog didžiausias potencialas,

atsižvelgiant į atliktus (giminingus) fonemų identifikavimo tyrimus, yra naudoti ITA arba DITA tipo tinklus.

Trečiame etape atliekamas gautų rezultatų palyginimas ir statistinių parametru skaičiavimas.



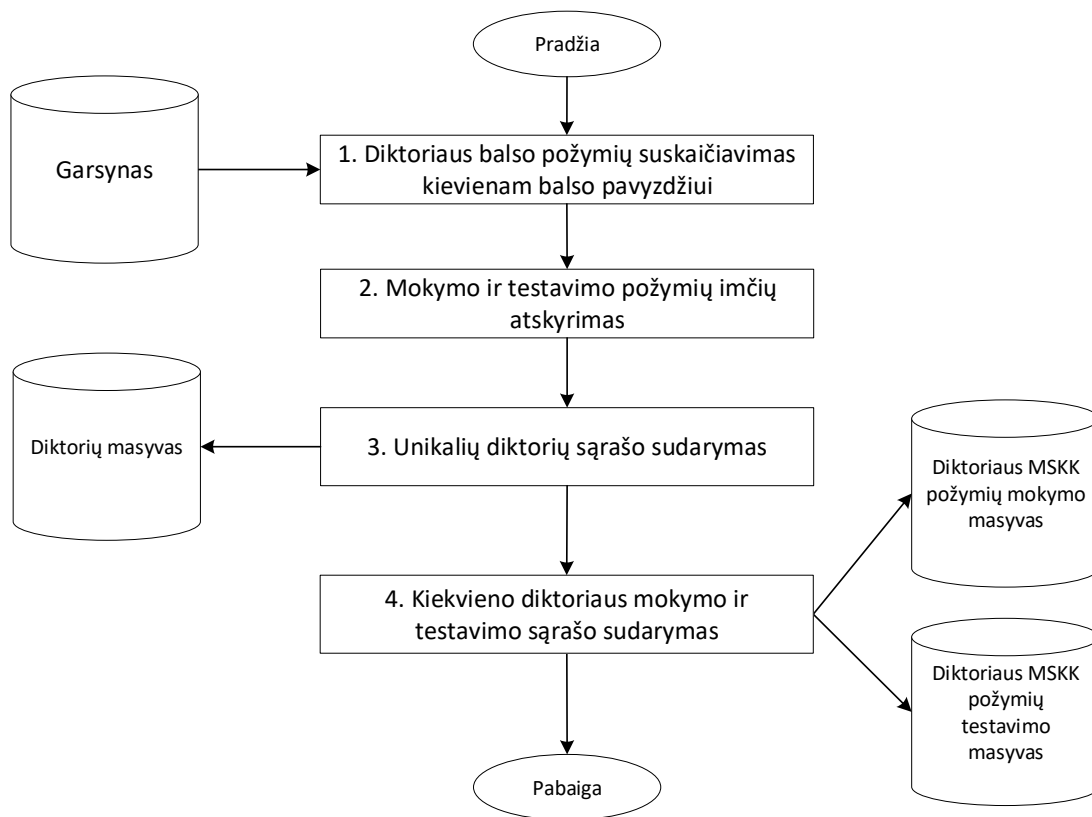
25 pav. Tyrimo veiksmų seka

2.2.1 Požymių išskyrimas

Dauguma autorių nurodo, jog naudoja standartu tapusius MSKK požymių rinkinius, aprašytus 1.2.2 skyriuje. Siekiant susiaurinti klasifikatorių tikslumo tobulinimo paiešką, buvo nuspręsta pasinaudoti fiksuotu parametru rinkiniu ir neieškoti papildomų požymių reikšmių, nes kitų autorių darbuose parodyta, kad toks rinkinys bene optimaliausias.

Šiam, lietuviškai kalbančio diktoriaus atpažinimo uždaviniui, buvo pasirinktas standartinis MSKK požymių rinkinys, kuris bus naudojamas tolimesniam modelio kūrimui bei apdorojimui. Naudojami 13 MSKK požymių, 13 delta, bei 13 delta-delta požymių su energijos koeficientais, iš viso – 39 pirminiai požymiai, gauti apdorojus balso signalo kadrą. Tyrimuose naudotas kadro dydis – 25ms, persidengimo riba – 10ms, su Hamingo lango funkcija.

Bendra požymių išskyrimo etapo blokinė diagrama pateikta 26-ame paveiksle.



26 pav. Melų skalės keistro koeficientų požymių išskyrimo etapų seka

1 žingsnis – diktoriaus balso požymių skaičiavimas kiekvienam balso pavyzdžiui. Pirmiausia reikia nustatyti turimo balso signalo parametrus, reikalingus MSKK skaičiavimui. Kaip buvo rašoma 1.2.2 skyriuje, balso įrašo signalas yra skaidomas į kadrus, tada iš jo išgaunami požymiai, kurie yra naudojami tolimesniam apdorojimui. Požymių išskyrimas aprašomas (Narayana ir Kopparapu, 2014) ir pateikiamas 1 algoritme. Jame yra parodyti požymių išskyrimo etapai, kuriuos reikia atlikti, kad kiekvienam diktoriaus balso signalo pavyzdžiui gautume MSKK požymių vektorių rinkinį.

1 algoritmas. Diktoriaus balso požymių išskyrimo pseudokodas

(Narayana ir Kopparapu, 2014)

Įėjimas:

$x[n]$ – kalbos signalas, kur $x_p[n]$ nurodo p -tajį kadra

f_s – semplavimo dažnis

P – kadrų skaičius

N – semplų ilgis

F – melo filtrų kiekis

$l_{f_{min}}$ – minimalus dažnis

$l_{f_{max}}$ – maksimalus dažnis, < nei Naikvisto dažnis signalui, bei < 6800 hz

Išėjimas:

$\Phi\{X\}$ – MSKK požymių matrica

ALGORITMAS:

$X \stackrel{\text{def}}{=} [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p, \dots, \vec{x}_P]$, X matricos dydis yra $N \times P$, kur $p = 1, 2, \dots, P$

#DFT skaičiavimas

$$X_p(k) = \sum_{n=0}^{N-1} x_p[n] w[n] e^{-\frac{j2\pi kn}{N}}, \quad 0 \leq k \leq N-1, \quad l_f(k) = kf_s/N,$$

#Hamingo langas

$$w[n] = 0.54 - 0.46 \cos\left(\frac{n\pi}{N}\right),$$

#melo filtrų banko skaičiavimas

$$M(m, k) = \begin{cases} 0, & \text{kai } l_f(k) < l_{f_c}(m-1) \\ \frac{l_f(k) - l_{f_c}(m-1)}{l_{f_c}(m) - l_{f_c}(m-1)}, & \text{kai } l_{f_c}(m-1) \leq l_f(k) < l_{f_c}(m) \\ \frac{l_{f_c}(m) - l_f(k)}{l_{f_c}(m) - l_{f_c}(m+1)}, & \text{kai } l_{f_c}(m) \leq l_f(k) < l_{f_c}(m+1) \\ 0, & \text{kai } l_f(k) \geq l_{f_c}(m+1) \end{cases},$$

#linijinių dažnių keitimas į melų skalę

$$\phi_f = 2595 * \log_{10}\left(1 + \frac{l_f}{700}\right),$$

#dažnių rezoliucija melų skalėje

$$\delta\phi_f = (\phi_{f_{max}} - \phi_{f_{min}})/(F + 1),$$

#centrinio dažnio melų skalėje skaičiavimas

$$\delta\phi_c = m * \delta\phi_f, \quad \text{kur } m = 1, 2, \dots, F$$

#melų keitimas į linijus dažnius

$$l_f(m) = 700 \left(10^{\frac{\phi_f(m)}{2595}} - 1\right)$$

#filtrų banko išėjimo logaritmas

$$L_p(m, k) = \ln\left\{\sum_{k=0}^{N-1} M(m, k) * |X_p(k)|\right\}, \quad \text{kur } m = 1, 2, \dots, F, \quad \text{kur } p = 1, 2, \dots, P$$

#DKT skaičiavimas

$$\Phi_p^r\{x[n]\} = \sum_{m=1}^F L_p(m, k) \cos\left\{\frac{r(2m-1)\pi}{2F}\right\}, \quad \text{kur } r = 1, 2, \dots, F \text{ ir } r\text{-tajį MFCC, } p\text{-tajam kadrai, } x[n] \text{ garso signalui}$$

#MSKK matrica

$$\Phi\{X\} = [\Phi_1, \Phi_2, \dots, \Phi_p, \dots, \Phi_P]$$

Return $\Phi\{X\}$

2 žingsnis – mokymo ir testavimo požymių imčių atskyrimas. Antrajame žingsnyje apskaičiuotus MSKK požymius kiekvienam diktoriui skirstome pagal mokymo ir testavimo imčių santykį. Atskyrimui visi diktorių požymių masyvai paskirstomi į dvi dalis: 70 proc. visų turimų pavyzdžių bus naudojama klasifikatoriaus apmokymui, likę 30 proc. – klasifikatoriaus testavimo tikslumui.

3 žingsnis – unikalių diktorių sąrašo sudarymas. Sąraše bus saugomi tik skirtingi diktoriai. Šiame etape reikia išrinkti unikalius diktorių pavadinimus iš visų turimų balso pavyzdžių. Šis sąrašas bus panaudotas atliekant klasifikatorių mokymą ir testavimą.

4 žingsnis – kiekvieno diktoriaus mokymo ir testavimo sąrašo sudarymas. Šiame žingsnyje reikia sudaryti mokymo ir testavimo pavyzdžių sąrašą kiekvienam diktoriui. Sąrašas kiekvienam diktoriui bus panaudotas atliekant klasifikatorių mokymą ir testavimą. Šis žingsnis reikalingas PMM klasifikatoriaus apmokymui.

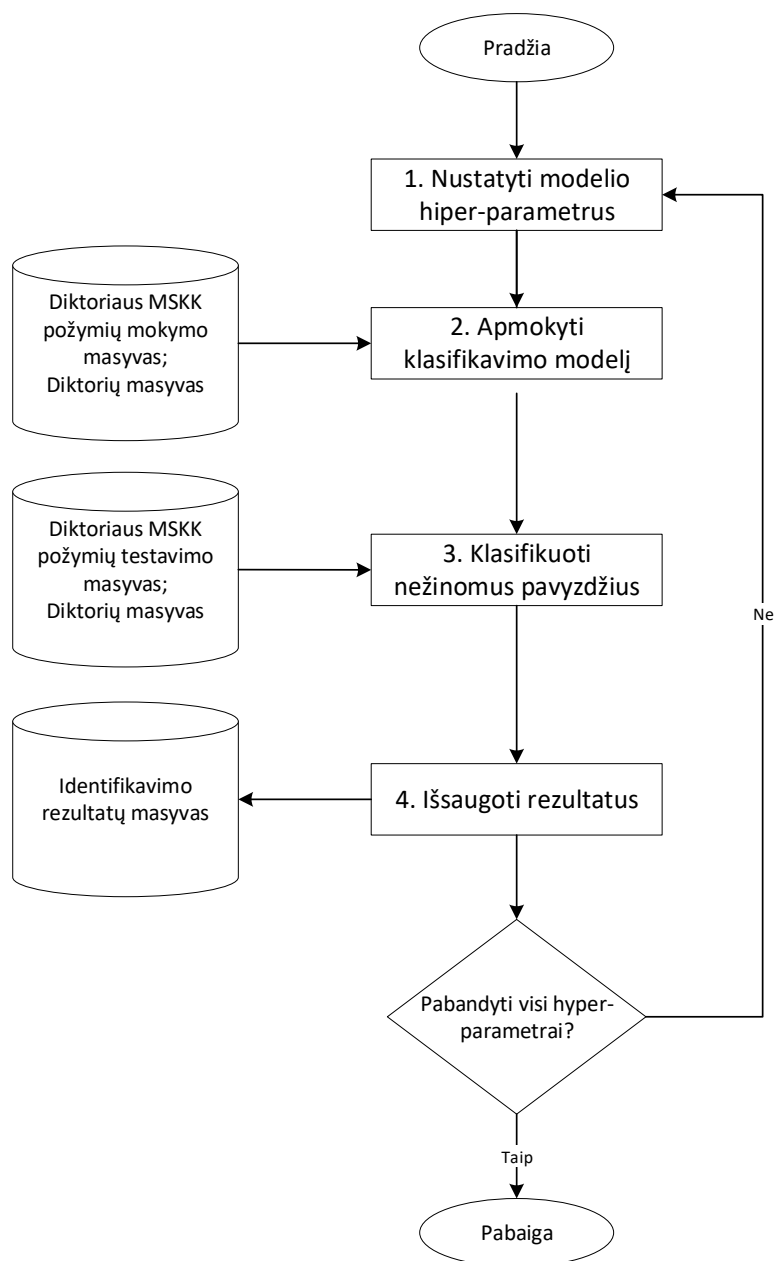
2.2.2 Klasifikatorių mokymas ir testavimas

Klasifikatorių mokymas atliekamas kaip veiksmų seka, parodyta 27-ame paveiksle. Schemoje išskirti šie pagrindiniai žingsniai:

1. Pagal iš anksto numatytas hiper-parametrų tikrinimo reikšmes, aprašytas 5-oje lentelėje, nustatomi klasifikavimo modelio hiper-parametrai.
2. Vykdomas klasifikatoriaus apmokymas, naudojant algoritmus, pateiktus 2.2.2.1 ir 2.2.2.2 skyriuose, kur aprašomi PMM ir RNT klasifikavimo modeliai. Mokymas vykdomas pagal 7-oje ir 8-oje lentelėse išskirtas mokymo imtis, kurios yra gautos po požymių etapo. Klasifikatoriaus apmokymas, tiek PMM, tiek RNT atveju, yra procesas, kurio metu yra nustatomos ir vėl perskaičiuojamos parametrų reikšmės, kol pasiekama didžiausia klasifikavimo tikslumo reikšmė arba pasiekama 200 iteracijų riba.
3. Trečiame žingsnyje paimami klasifikatoriaus apmokymui nenaudoti testavimo pavyzdžiai. Jie panaudojami klasifikatoriaus tikslumo įvertinimui nustatyti.

4. Gauti testavimo pavyzdžių atpažinimo rezultatai, klasifikatoriaus atpažintas diktorius ir tikroji diktorius reikšmė yra išsaugomi ir naudojami vėliau, išrenkant geriausią klasifikatoriaus variantą su garsyno diktorių pavyzdžiais.

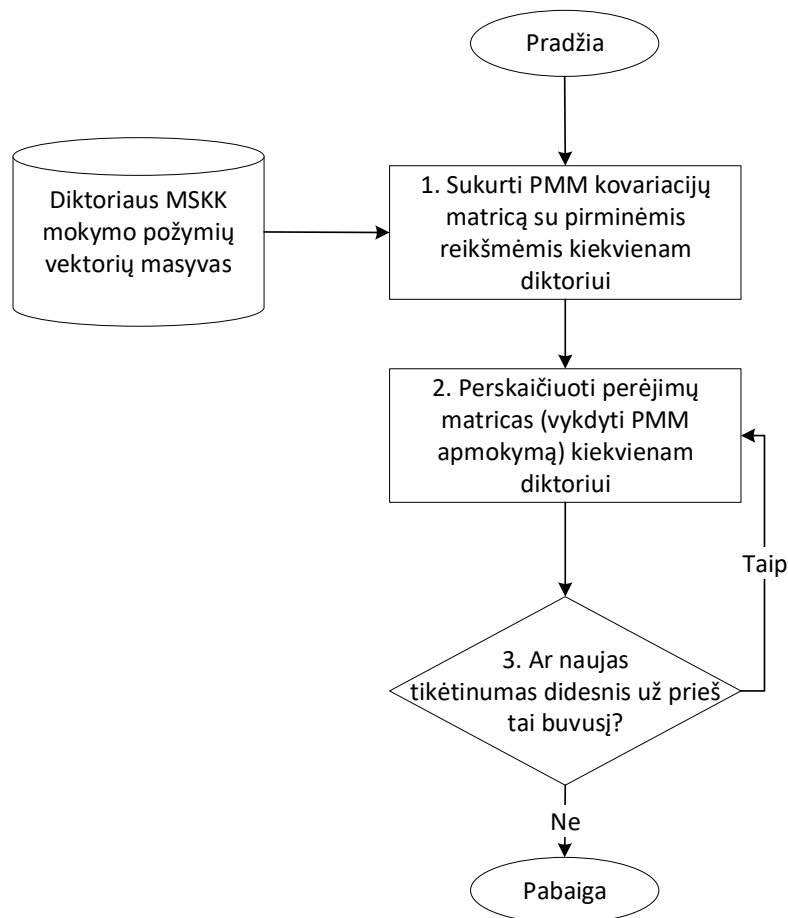
Ši veiksmų seka yra kartojama tol, kol išbandomi visi klasifikatorių hiper-parametrai: neuroninio tinklo topologija, gylis, neuronų skaičius, neuronų atjungimo parametras. Konkretūs parametrai pateikiami 5-oje lentelėje.



27 pav. Klasifikatorių hiper-parametrų paieškos ciklas

2.2.2.1 Paslėptųjų Markovo modelių klasifikatorius

Paslėptųjų Markovo modelių seką reikia apmokyti, o diktoriaus identifikavimo tikslumas priklauso nuo PMM būsenų skaičiaus ir būsenų išėjimų tikimybių reikšmių ir perėjimų tarp būsenų tikimybių reikšmių. Pirminis būsenų kiekis buvo pasirenkamas įvertinant vidutinę įrašo trukmę, o būsenų perėjimų matricos užpildomos pseudo-atsitiktiniais koeficientų rinkiniais. PMM modelio parinkimas vyko pagal 28-ame paveiksle atvaizduotą veiksmų seką.



28 pav. Paslėpto Markovo modelio apmokymo veiksmų seka

Šioje diagramoje (28 pav.) išskiriami 3 žingsniai:

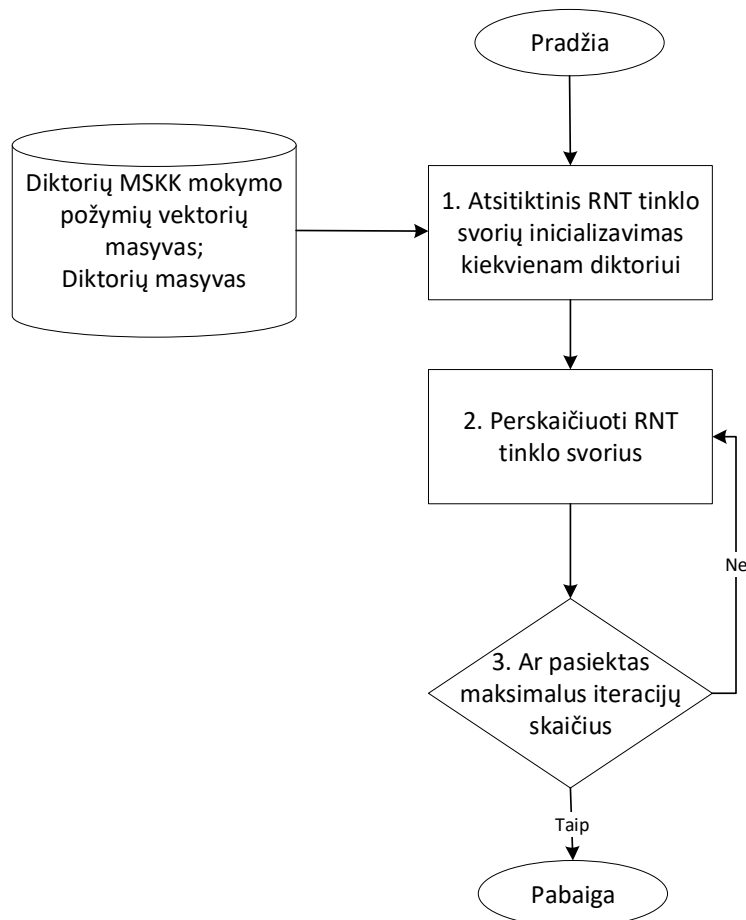
1. Kiekvienam diktoriui sudaroma pirminė (preliminari) PMM būsenų perėjimų matrica atsitiktinai;
2. Perskaičiuojama diktoriaus PMM būsenų perėjimų tikimybių matrica, naudojant Baum-Welch algoritmą (Duda, Hart ir Stork 2000);
3. Palyginama, ar prieš 3 žingsnį išsaugota būsenų perėjimų tikimybių matrica turėjo didesnę diktoriaus atpažinimo tikėtinumą, lyginant su naujai

perskaičiuota. Jei taip, naujai perskaičiuota matrica išsaugoma ir kartojamas 2 žingsnis, jei ne – apmokymas baigiamas.

Aprašytuose žingsniuose PMM apmokomas naudojant (Rabiner 1989) aprašytą mokymo algoritmą.

2.2.2.2 Rekurentinių neuroninių tinklų klasifikatorius

Atliekant neuroninio tinklo mokymą, pagrindinis skirtumas, lyginant su paslėptų Markovo modelių mokymu, yra tas, kad neuroninio tinklo svoriai yra keičiami pagal kiekvieną diktoriaus mokymo pavyzdį visam neuroniniam tinklui. Taip pat mokymas yra stabdomas, pasiekus maksimalų mokymo iteracijų skaičių. Tuo tarpu atliekant PMM mokymą, būsenų perėjimo tikimybių matricos yra tikslinamos kiekvienam diktoriui atskirai. RNT neuroninio tinklo mokymo veiksmų seka atvaizduota 29-ame paveiksle.



29 pav. Rekurentinio neuroninio tinklo apmokymo veiksmų seka

Rekurentinio neuroninio tinklo apmokymo bendroje veiksmų sekoje įvardyti šie žingsniai:

1. Neuroniniam tinklui yra pateikiama visų diktorių mokymo pavyzdžių MSKK požymių aibė su atitinkamai sužymėtais unikaliais diktorių pavadinimais. Pirminiai tinklo svoriai pasirenkami generuojant atsitiktines reikšmes.
2. Pateiktiems pavyzdžiams apskaičiuojami tinklo išėjimai, įvertinama klaida, toliau svoriai perskaičiuojami, siekiant minimizuoti klaidos funkciją.
3. Priimamas sprendimas, tęsti ar nutraukti mokymo veiksmą, jei yra pasiekta 200 mokymo iteracijų riba.

Pateikta seka naudojama ieškant tiksliausią diktoriaus identifikavimą atliekančio neuroninio tinklo topologijos. Tikslas – rasti tokius svorių koeficientus, prie kurių klasifikatorius tiksliausiai identifikuotų diktoriaus klasę pagal testavimo imties diktoriaus MSKK požymius.

2.2.3 Rezultatų analizė ir palyginimas

Siekiant įvertinti pasiūlytų klasifikatorių tikslumą, sudaryti palyginamieji rezultatų masyvai. Naudojant šiuos masyvus, galima ne tik įvertinti identifikavimo tikslumą konkrečiam klasifikatoriui, bet ir apskaičiuoti papildomą statistiką, pavyzdžiui, rezultatų patikimumo intervalus. Klasifikatorių palyginimas gali būti atliktas keletu metodu (Dietterich, 1998), tarp kurių – McNemar testas ir kryžminė validacija (angl. *cross validation*) (Arlot ir Celisse, 2010). T. G. Dietterich (1998) klasifikatorių palyginimui rekomenduoja turint neribotus skaičiavimo resursus naudoti kryžminę validaciją, o esant apribojimams – McNemar testą.

Visa detalios rezultatų analizės eiga atvaizduota 30-ame paveiksle.

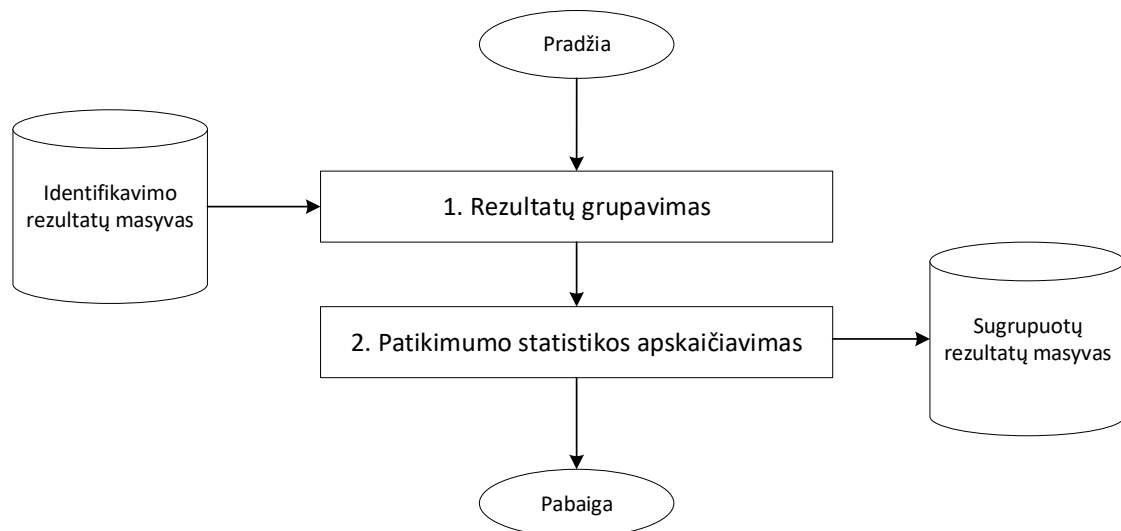
1 žingsnis – rezultatų grupavimas. Pirmiausia sukuriamas visų rezultatų masyvas, kuriame yra unikalaus pavyzdžio numeris ir kiekvieno klasifikavimo būdo klasifikavimo reikšmė. Toliau pagal algoritmą priskiriamas klaidos kodas,

kur kiekvienam atpažinimo atvejui priskiriamas vienas iš galimų rezultato žymėjimų:

- #1 – PMM ir RNT identifikavimo rezultatas atitinka tikrąjį diktorių
- #2 – PMM ir RNT neteisingai identifikavo diktorių
- #3 – PMM identifikavo teisingai, RNT klydo
- #4 – RNT identifikavo teisingai, PMM klaidingai

Šie rezultatų žymėjimai yra naudojami, kad būtų surinkta statistika apie atpažinimo tikslumo statistinį reikšmingumą.

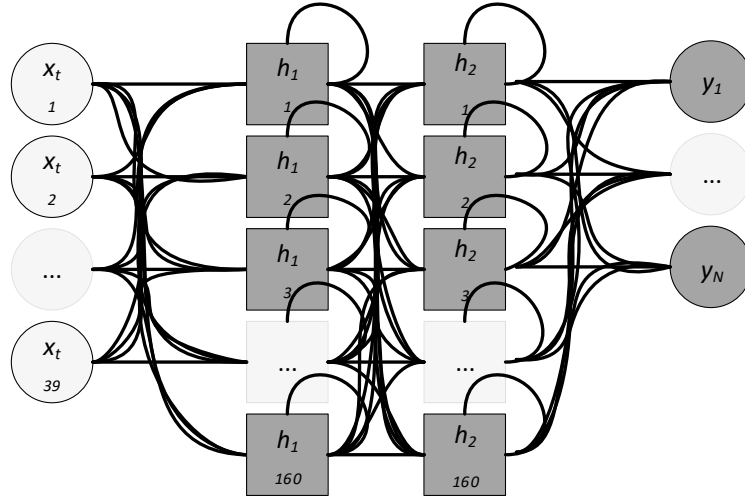
2 žingsnis – patikimumo statistikos apskaičiavimas. Siekianti įvertinti skirtingų klasifikatorių patikimumą, yra skaičiuojama McNemar testo statistika (Dietterich, 1998), kuri šiuo atveju parodo, ar diktoriaus identifikavimo tikslumo skirtumas tarp PMM ir RNT klasifikatorių yra statistiškai reikšmingas.



30 pav. Detalios rezultatų analizės diagrama

2.3 Siūlomas diktoriaus identifikavimo metodas

Siūlomas diktoriaus identifikavimo metodas naudoja rekurentinio neuroninio tinklo struktūrą, sudarytą iš dviejų sluoksnių po 160 neuronų ir išėjimo sluoksnio, kuris yra visiškai sujungtas ir turi *softmax* aktyvaciją, kaip pavaizduota 31-ame paveiksle.



31 pav. Siūloma neuroninio tinklo topologija

Neuroninio tinklo įėjimai $x_{t1}, x_{t2}, \dots, x_{t39}$ sudaryti iš 39 dimensijų MSKK požymių vektoriaus su laiko seka t žingsniu nuo 1 iki 999, gaunamu iš kiekvieno diktoriaus balso įrašo, apskaičiuojamo pagal 1 algoritmą ir praplėstas delta bei delta-delta koeficientais, kurie detalčiau aprašomi 2.2.1 skyriuje.

Tinklo neuronai $h_{1-1}, h_{1-2}, \dots, h_{2-160}$ – tai ITA tinklo celės (Graves, Mohamed, ir Hinton, 2013), apskaičiuojamos pagal formulę (32) su įėjimo loginiu elementu (i_t) pagal formulę (28), užmiršimo loginiu elementu (f_t) pagal formulę (29), celės loginiu elementu (c_t) pagal formulę (30) ir išėjimo loginiu elementu (o_t) (31).

Tinklo išėjimo sluoksnis $y_1, \dots, y_N$ – tai pilnai sujungtas neuroninio tinklo sluoksnis su *softmax* aktyvacija, pagal formulę (19), kur N – visų diktorių skaičius.

Siūlomo neuroninio tinklo apmokymas. Prieš šio neuroninio tinklo panaudojimą diktoriaus identifikavimui, jį reikia apmokyti. Tinklo apmokymas vykdomas pradinis paslėptų sluoksnių įėjimo celės svorius W_{xi}, W_{hi}, W_{ci} inicializuojant pagal X. Glorot ir Y. Bengio (2010) pasiūlytą metodą. Vidiniai celių svoriai $W_{xf}, W_{hf}, W_{cf}, W_{xc}, W_{hc}, W_{cf}, W_{xo}, W_{ho}, W_{co}$ inicializuojami pagal A. M. Saxe, J. L. McClelland ir S. Ganguli (2014) metodą.

ITA tinklo celių dvikryptiškumas yra įgyvendinamas pritaikant RNT dvikryptiškumo principą pagal M. Schuster ir K. K. Paliwal (1997) metodą, kai

neuroninio tinklo mokymui yra taikoma dvikryptiškumo procedūra, plačiau aprašoma 1.4.2.2 skyriuje.

Tinklo mokymui taikoma retos kryžminės entropijos (angl. *sparse categorical cross entropy*) nuostolio funkcija (De Boer ir kt., 2005) su Adam optimizacija pagal D. P. Kingma ir J. L. Ba (2015).

2.4 Skyriaus išvados

Šiame skyriuje pristatyti siūlomi metodai diktoriaus identifikavimo tikslumo tyrimui ir lietuviškai kalbančio diktoriaus identifikavimo metodas. Diktoriaus identifikavimo tikslumui įvertinti pirmiausia naudojamas požymių išskyrimo algoritmas, naudojant MSKK požymius iš diktoriaus balso signalo įrašo. Toliau, apmokant diktoriaus identifikavimo PMM ir RNT klasifikatorius, naudojami aprašyti mokymo algoritmai.

Apžvelgtas siūlomas diktoriaus identifikavimo metodas:

- Lietuviškai kalbančio diktoriaus identifikavimas, naudojant ilgos trumpalaikės atminties rekurentinio neuroninio tinklo celes. Šio metodo diktoriaus identifikavimo tikslumą pravartu išbandyti eksperimentiškai, nes metode pasiūlyti naudoti neuroniniai tinklai giminingose šnekos atpažinimo srityse parodė tikslesnius identifikavimo rezultatus, nei klasikiniai identifikavimo metodai.

3 EKSPERIMENTINIO TYRIMO REZULTATAI

Šio skyriaus pirmoje dalyje yra detalai aprašomas eksperimentiniams tyrimams naudotas garsynas ir jo techniniai parametrai. Toliau skyriuje yra pateikiami identifikavimo tikslumo tyrimų rezultatai ir šių rezultatų statistinis reikšmingumas.

Tarpiniai skyriuje pateikiamų klasifikatorių identifikavimo tikslumo rezultatai buvo publikuoti moksliniame žurnale (Dovydaitis ir Rudžionis 2018).

3.1 Garsynas LIEPA

Lietuvių šnekamosios kalbos garsynas LIEPA – fonetiškai reprezentatyvi lietuvių šnekos duomenų bazė, pritaikyta elektroninių paslaugų kūrimui bei šnekos technologijų moksliniams tyrimams ir konstravimo darbams (Laurinčiukaitė ir kt. 2017). Garsyną sudaro šnekos atpažinimo paslaugų kūrimo tikslams surinkta 100 val. trukmės (garsyno dalis Nr. 1) ir šnekos sintezavimo tikslams surinkta 13 val. trukmės (garsyno dalis Nr. 2) dalys.

Garsyne LIEPA yra naudojami 78 skirtingų tekstų įrašai, 45 iš jų apima nepertraukiamos rišlios kalbos įrašų pavyzdžius. Įrašai buvo renkami daugiau nei 6 vietose su skirtinga įranga.

Renkant pirmą garsyno dalį, iš kurios yra sudarytos visos šiame tyrime naudotos mokymo ir testavimo pavyzdžių imtys, iš viso dalyvavo 376 diktoriai – 116 mokinių, 260 universiteto studentų. Šį skaičių sudarė 248 moterys ir 128 vyrai.

Pagrindinės techninės garsyno charakteristikos yra pateikiamos 6-oje lentelėje. Čia nurodoma garsyno trukmė valandomis, šnekos turinio aprašymas, bendras diktorių kiekis, diskretizavimo dažnis, kvantavimo rezoliucija ir kanalų kiekis.

6 lentelė. Garsyno LIEPA charakteristikos

S. Laurinčiukaitė ir kt. (2017)

Kriterijus	Garsyno charakteristika
Garsyno apimtis	100 valandų
Šnekos turinys	Žodžiai, frazės, sakiniai, tekstai
Diktorių kiekis	376
Diskretizavimo dažnis	22kHz
Kvantavimo rezoliucija	16 bitų
Kanalų kiekis	1

Kaip teigia S. Laurinčiukaitė ir kt. (2017), LIEPA garsynas gali būti sėkmingai naudojamas tiek įvairių atpažinimo sistemų tyrimui, tiek praktinių taikymų kūrimui.

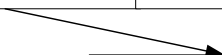
Garsyno adaptavimas diktoriaus identifikavimo tikslumo tyrimui.

Prieš pradėdant tyrimą ir siekiant efektyviai panaudoti turimus programinės ir aparatinės įrangos resursus, tyrimų eiga buvo suskirstyta į du etapus:

1. Bandomąjį (pilotinį) etapą, kuriam išskirta tik dalis visos LIEPA garsyno diktorių bazės;
2. Galutinį rezultatų patikrinimą, testavimui naudojant visų LIEPA garsyne esančių diktorių pavyzdžius.

Pirminiai bandymai buvo atlikti naudojant atskirą diktorių duomenų rinkinį, kad būtų patikrintas programinės įrangos veikimo teisingumas ir įvertintos pasirinktų klasifikavimo metodų galimybės identifikuoti lietuviškai kalbančius diktorius maksimaliai tiksliai. Šis išskirstymas buvo atliktas darant prielaidą, kad klasifikatoriaus identifikavimo tikslumas bus panašus, kadangi garsyną sudarantys diktorių įrašai yra surinkti panašiomis sąlygomis. Pilnas eksperimento tyrimas buvo išplėstas naudojant LIEPA įrašų duomenų bazę, visus diktorių pavyzdžius padalinus į 5 dalis.

Bandomųjų tyrimų imtis. Prieš atliekant testus naudojant visą LIEPA garsyną, pirminiams diktoriaus identifikavimo bandymams garsynas padalintas į skirtingas dalis, kad būtų atlikti bandomieji skaičiavimai ir pilnos apimties eksperimentas. Garsyno suskirstymas pateikiamas 32-ame paveiksle.

Garsynas LIEPA. Visi diktorių balso pavyzdžiai	Nepasikartojantis pirmas diktorių pavyzdžių rinkinys	Mokymo imtis
		Testavimo imtis
	Nepasikartojantis antras diktorių pavyzdžių rinkinys	Mokymo imtis
		Testavimo imtis
	Nepasikartojantis trečias diktorių pavyzdžių rinkinys	Mokymo imtis
		Testavimo imtis
	Nepasikartojantis ketvirtas diktorių pavyzdžių rinkinys	Mokymo imtis
		Testavimo imtis
	Nepasikartojantis penktas diktorių pavyzdžių rinkinys	Mokymo imtis
		Testavimo imtis
		
Atsitiktinai parinktas 66 diktorių balso pavyzdžių rinkinys bandomiesiems tyrimams	Mokymo imtis	
	Testavimo imtis	

32 pav. Garsyno suskirstymas į bandomąją imtį ir 5 dalis pilnam eksperimentui

Be anksčiau minėtų priežasčių, bandomojo garsyno imties išskyrimas (7 lentelė) suteikia galimybę išbandyti daugiau skirtingų klasifikatorių variantų, nes diktorius akustinis modeliavimas naudoja gana nemažai skaičiavimo resursų. Tai taip pat leidžia sutrumpinti tyrimų laiką.

7 lentelė. Garsyno LIEPA bandomojo eksperimento charakteristika

Parametro pavadinimas	Parametro reikšmė
Unikalių diktorių skaičius	66
Bendra pavyzdžių imtis	4691
Mokymo imtis	3317
Testavimo imtis	1374

Pilnos apimties garsynas. Atrinkus aukščiausią tikslumą teikiančius parametrus ir iš jų išrinkus tiksliausiai diktorių identifikuojančių klasifikatorių parametrus, buvo pereita prie antros eksperimento dalies – darbo su pilna pavyzdžių baze, kad būtų patikrinti bandomųjų tyrimų metu išrinkti tiksliausi identifikavimo metodai. Kadangi bandomasis tyrimas buvo darytas su daugiau nei 60 unikalių diktorių, pasirinkta, kad LIEPA garsynas turi būti suskaidytas į 5-ias dalis (8-ta lentelė), kur kiekvienos dalies kiekviename rinkinyje yra ne mažiau kaip 60 unikalių diktorių įrašų.

8 lentelė. Garsyno LIEPA pilno eksperimento rinkinių charakteristika

Parametro pavadinimas	Pirmas rinkinys	Antras rinkinys	Trečias rinkinys	Ketvirtas rinkinys	Penktas rinkinys
Diktorių skaičius	61	62	62	60	61
Pavyzdžių imtis	4408	3523	5528	6090	4155
Mokymo imtis	3119	2497	3905	4292	2934
Testavimo imtis	1289	1026	1623	1798	1221

Modifikuotas garsynas LIEPA su papildomu triukšmu. Siekiant palyginti pasiūlyto metodo efektyvumą, pilnos apimties garsynas buvo modifikuotas, jį aptriukšminant. Aptriukšminimui buvo naudotas baltas triukšmas (angl. *white noise*) su 2 proc. dydžio amplitude. Visi kiti garsyno parametrai išlaikyti tokie patys, kaip ir pradiniam garsyne.

3.2 Tyrimai naudojant geriausią klasifikatorių

Šioje dalyje yra atliekami geriausio klasifikatoriaus tyrimai pagal 2 skyriuje pristatytą tyrimo metodiką.

3.2.1 Tyrimas su bandomąja garsyno imtimi

Pirminiai tyrimai atlikti su bandomosios imties pavyzdžiais, kurie yra išskirti 7-oje lentelėje. Su šia pavyzdžių imtimi atlikti planuoti eksperimentai pagal iš anksto numatytus parametrų rinkinius 8-oje lentelėje pateikiamais klasifikatorių hiper-parametrais.

PMM bandomosios imties tyrimas. Atlikus paslėptųjų Markovo modelių apmokymą diktoriaus identifikavimui, buvo atlikti tyrimai, siekiant rasti suboptimalius šiam uždaviniui hiper-parametrus. Paieška atlikta paprasto perrinkimo būdu. Gauti rezultatai pavaizduoti 9-oje lentelėje. Lentelėje pateikiami gauti diktoriaus identifikavimo rezultatai, naudojant PMM ir keičiant būsenų skaičių.

9 lentelė. PMM tinklo atpažinimo rezultatų priklausomybė nuo PMM būsenų skaičiaus

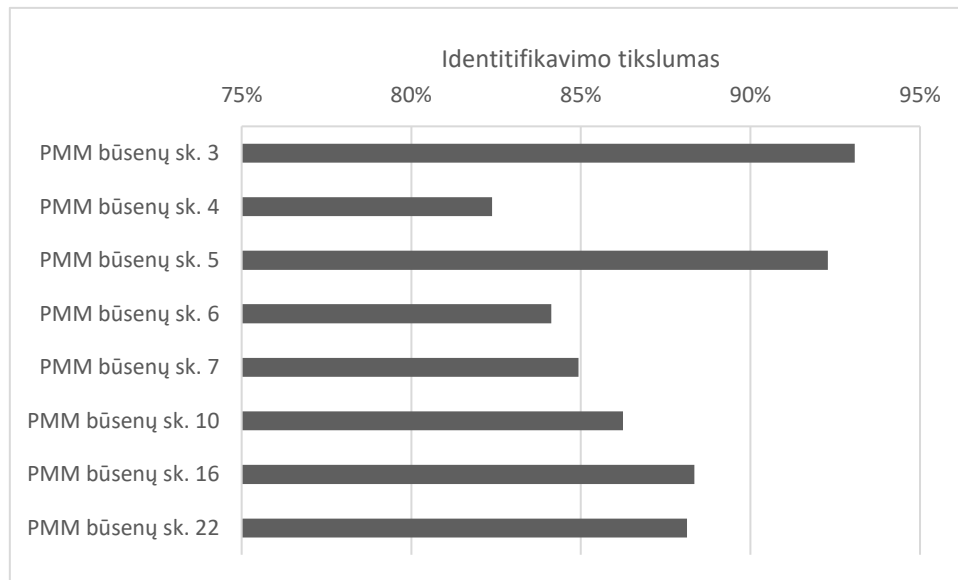
Modelio būsenų skaičius	Identifikavimo tikslumo geriausias rezultatas
5	92,28 proc.
7	84,93 proc.
10	86,24 proc.
16	88,35 proc.
22	88,13 proc.

Atlikus pirminius bandymus, gauti rezultatai leido įvertinti būsenų skaičiaus tikslinimo režius. Papildomas būsenų skaičiaus įtakos įvertinimas atliktas aplink tiksliausiai identifikavusį PMM būsenų skaičių. Gauti rezultatai pateikiami 10-oje lentelėje.

10 lentelė. PMM tinklo patikslinto tyrimo rezultatai su bandomąja imtimi

Modelio būsenų skaičius	Identifikavimo tikslumo geriausias rezultatas
3	93,08 proc.
4	82,38 proc.
6	84,13 proc.

Vizualizavus gautus rezultatus ir juos apibendrinus, darome išvadą, kad aukščiausias diktoriaus identifikavimo tikslumas gaunamas, kuomet būsenų skaičius yra santykinai nedidelis, o ryškus būsenų skaičiaus didinimas neduoda identifikavimo tikslumo pagerėjimo (33 paveikslas).



33 pav. Bandomosios imties identifikavimo tikslumo grafikas, naudojant paslėptus Markovo modelius

Iš pateikto grafiko matyti, jog du didžiausią tikslumą gaunantys atvejai yra tuomet, kai PMM paslėptų būsenų skaičius yra 3 ir 5. Todėl vykdant pilnos apimties eksperimentą, bus naudojamas 3 būsenų skaičius, leidęs pasiekti geriausius rezultatus pilotinio tyrimo metu. 3 PMM būsenos yra minimalus galimas būsenų skaičius. Numanoma, kad per didelis būsenų skaičius pats tampa triukšmu, neleidžiančiu tiksliau identifikuoti diktoriaus.

RNT bandomosios imties tyrimas. Kaip ir PMM bandomosios imties tyrimo atveju, pagal iš anksto numatytus parametrus atliekame pirminius ir patikslinamuosius tyrimus, nustatant suboptimalius RNT tinklų hiperparametrus, naudojant 5-oje lentelėje išvardintus parametrus. Kadangi RNT tyrimas suskaidytas į tris skirtingus etapus, pradedame nuo etapo, kurio metu sieksime nustatyti tiksliau diktorių identifikuojančią RNT topologiją. Pirmojo etapo tyrimo rezultatai pateikti 11-oje lentelėje.

11 lentelė. RNT tinklo atpažinimo rezultatų priklausomumas nuo tinklo topologijos

Tinklo topologija ir neuronų skaičius	Identifikavimo tikslumo geriausias rezultatas
ITA 80	70,38 proc.
DITA 80	83,77 proc.

Kaip matome iš pirmojo RNT tyrimo etapo, tikslesnis identifikavimo rezultatas yra gaunamas su DITA tipo topologijos tinklu. Taip pat pastebime, kad pakeitus pradinę tinklo topologiją ir vykdant apmokymą naudojant dvikryptės topologijos variantą, diktoriaus identifikavimo tikslumas pastebimai padidėja.

Antrojo patikslinamojo RNT etapo tikslas – nustatyti patobulintą tinklo topologiją, pritaikant atsitiktinį neuronų atjungimą. Tyrimas atliktas dviem etapais. Pirmojo eksperimento metu nustatytas suboptimalus atjungimo sluoksnio režis tarp 0,2, 0,4 ir 0,6 koeficientų, o patikslinimo metu iširti gretimi atjungimo koeficientai (12-ta lentelė).

12 lentelė. RNT tinklo atpažinimo rezultatai, įvertinant skirtingus atjungimo sluoksnio koeficientus

Tinklo topologija	Neuronų atjungimo koeficientas	Identifikavimo tikslumo geriausias rezultatas
DITA 80	0,2	90,90 proc.
DITA 80	0,3	90,10 proc.
DITA 80	0,4	87,04 proc.
DITA 80	0,5	80,42 proc.
DITA 80	0,6	71,47 proc.

Iš gaunamų rezultatų matome, kad tiksliausias identifikavimo rezultatas gaunamas, kai neuronų atjungimo koeficientas svyruoja tarp 0,2 ir 0,3. Didinant neuronų atjungimo koeficientą, identifikavimo tikslumas ima prastėti.

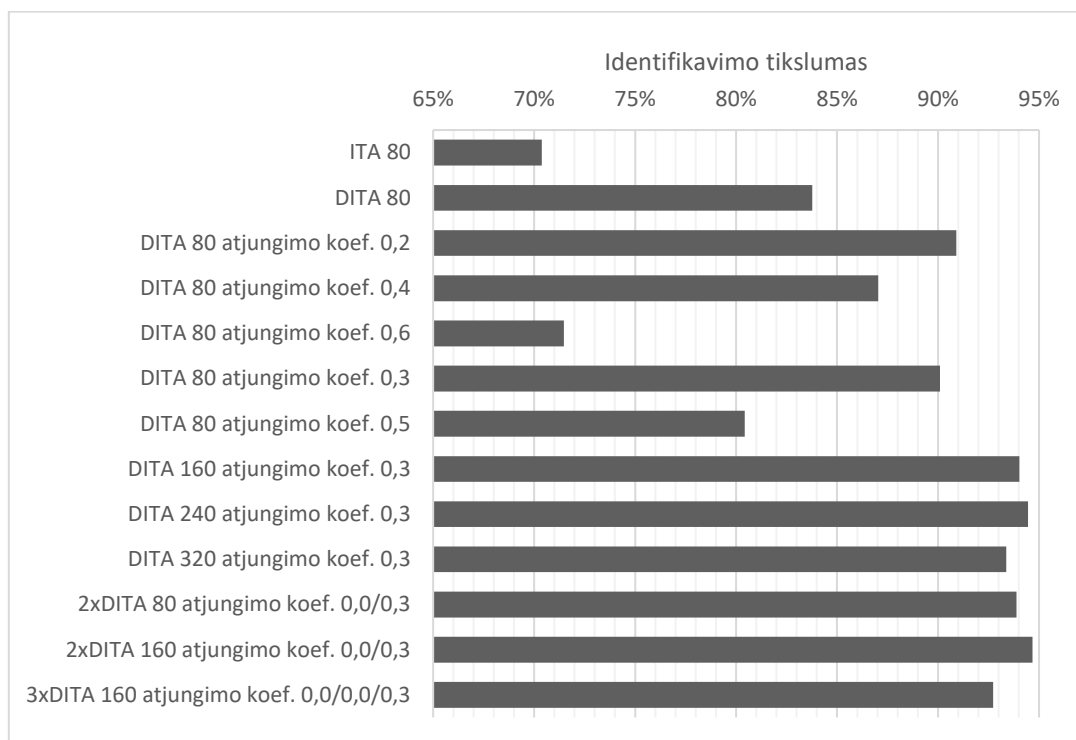
Trečiajame neuroninio tinklo hiper-parametrų tyrimo etape siekta išsiaiškinti neuroninio tinklo pločio ir gylio įtaką sprendžiamo uždavinio rezultatams. Pirmame žingsnyje tiriama skirtingo tinklo pločio įtaka tikslumui (13-ta lentelė), antrame – gylio.

13 lentelė. RNT tinklo atpažinimo rezultatai, patikslinus atjungimo sluoksnio koeficientus

Tinklo topologija	RNT plotis	Identifikavimo tikslumo geriausias rezultatas
DITA atjungimo koef. 0,3	160	94,03 proc.
DITA atjungimo koef. 0,3	240	94,46 proc.
DITA atjungimo koef. 0,3	320	93,37 proc.
DITA atjungimo koef. 0,0/0,3	80x80 (x2)	93,88 proc.
DITA atjungimo koef. 0,0/0,3	160x160 (x2)	94,68 proc.
DITA atjungimo koef. 0,0/0,0/0,3	160x160x160 (x3)	92,72 proc.

Lyginant padidintą neuronų skaičių su tinklo gylio padidiniu, yra stebimas nedidelis identifikavimo tikslumo skirtumas, kuris, atlikus pirmąjį padidinimą, abiem atvejais padidina klasifikavimo tikslumą. Jei tinklo plotis ar gylis toliau didinamas, klasifikavimo rezultatai tampa prastesni.

Apibendrinant atliktus bandomuosius tyrimus su RNT konfigūracijomis, pateikiamas grafikas (34-as paveikslas), kuriame parodyta diktorius identifikavimo tikslumo priklausomybė nuo tinklo tipo, neuronų skaičiaus ir atjungimo koeficiento.

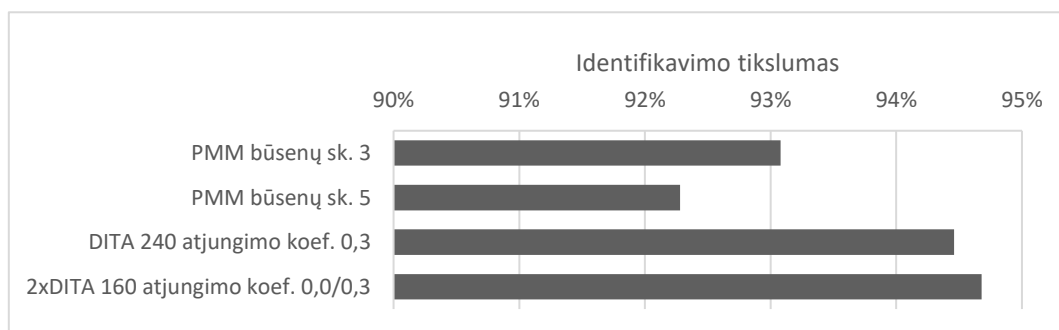


34 pav. Bandomosios imties identifikavimo tikslumo grafikas, klasifikavimui naudojant rekurentinius neuroninius tinklus

Iš pateiktų rezultatų matyti, jog aukščiausią diktorius identifikavimo tikslumą pasiekę klasifikatoriai yra tokių topologijų tinklai: DITA 240 su atjungimo koef. 0,3 ir DITA 2x160 su atjungimo koef. 0,0/0,3.

Bendras bandomosios imties pavyzdžių klasifikavimo palyginimas.

35-ame paveiksle palyginami du geriausi PMM (būsenų skaičius 3 ir 5) ir RNT (DITA 240 su atjungimo koef. 0,3 ir DITA 2x160 su atjungimo koef. 0,0/0,3) modeliai, kurie bus toliau testuojami su pilnos imties diktorių baze.



35 pav. Bandomosios imties tiksliausių rezultatų palyginimo grafikas

3.2.2 Tyrimai naudojant pilną LIEPA garsyną

Tyrimai naudojant pilnos apimties duomenų bazę bus atlikti su 4 geriausius rezultatus pirminio tyrimo metu leidusiais pasiekti modeliais ir jų hiperparametrais.

PMM atveju nustatyta, kad diktoriaus identifikavimas atliekamas tiksliausiai, kai paslėptų būsenų skaičius buvo:

- 3;
- 5.

RNT atveju, nustatyta, kad tiksliausią identifikavimą atlikusios tinklų konfigūracijos buvo:

- DITA 240 atjungimo koef. 0,3;
- DITA 2x160 atjungimo koef. 0,0/0,3.

14-oje lentelėje matomas PMM ir RNT identifikavimo tikslumas visai pavyzdžių imčiai.

14 lentelė. PMM ir RNT geriausių klasifikatorių rezultatai

Klasifikatorius	Pirmas rinkinys	Antras rinkinys	Trečias rinkinys	Ketvirtas rinkinys	Penktas rinkinys
PMM 3 būsenos	89,29 proc.	82,35 proc.	84,53 proc.	93,10 proc.	90,00 proc.
PMM 5 būsenos	70,75 proc.	60,62 proc.	62,23 proc.	89,76 proc.	77,23 proc.
DITA 240 atjungimo koef. 0,3	90,30 proc.	83,91 proc.	92,48 proc.	97,66 proc.	95,82 proc.
2xDITA 160 atjungimo koef. 0,0/0,3	92,08 proc.	86,45 proc.	92,35 proc.	97,10 proc.	96,39 proc.

Pasvertasis tikslumo vidurkis šiose lentelėse parodo visų tyrimo grupių identifikavimo tikslumo koeficientą, jį skaičiuojant pagal kiekvienos grupės

testavimo pavyzdžių skaičių. Pasvertasis vidurkis (PV) skaičiuojamas pagal formulę:

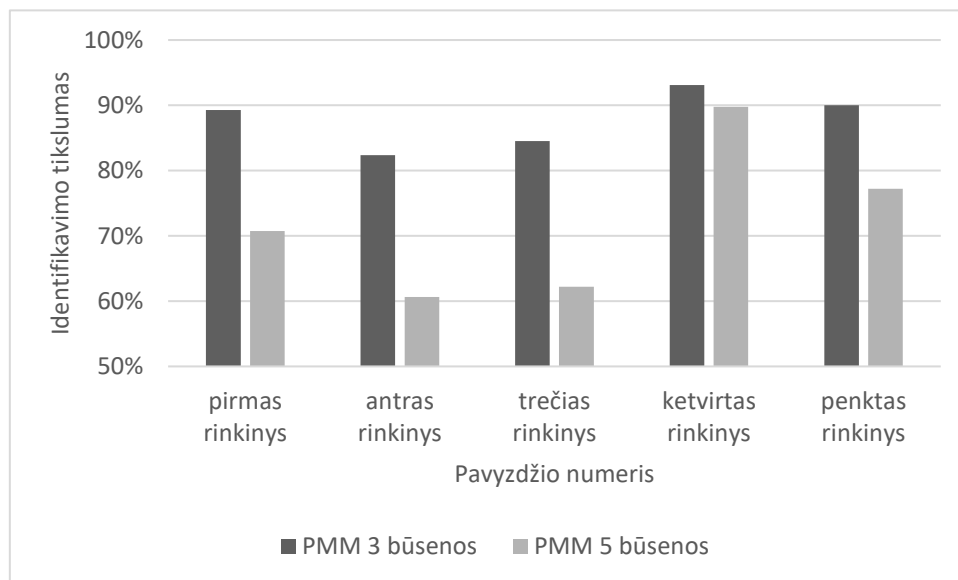
$$PV = \frac{\sum_{n=1}^N (E_n Z_n)}{\sum_{n=1}^N Z_n}, \quad (33)$$

kur E – tikslumas n -ajam rinkiniui, Z – testavimo pavydžių imtis rinkinyje, N – rinkinių kiekis.

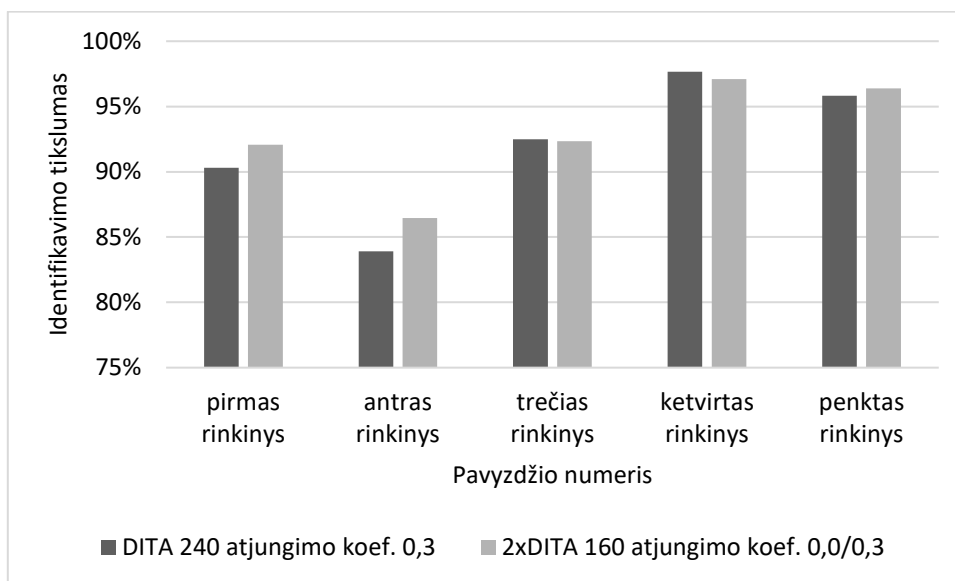
15 lentelė. PMM ir RNT geriausių klasifikatorių rezultatų vidurkiai visiems rinkiniams

Klasifikatorius	Tikslumo vidurkis	Pasvertas tikslumo vidurkis
PMM 3 būsenos	87,85 proc.	88,27 proc.
PMM 5 būsenos	72,12 proc.	73,32 proc.
DITA 240 atjungimo koef. 0,3	92,03 proc.	92,74 proc.
2xDITA 160 atjungimo koef. 0,0/0,3	92,87 proc.	93,37 proc.

Toliau pateikiami identifikavimo tikslumo palyginamieji grafikai (36 ir 37-ame paveiksluose).

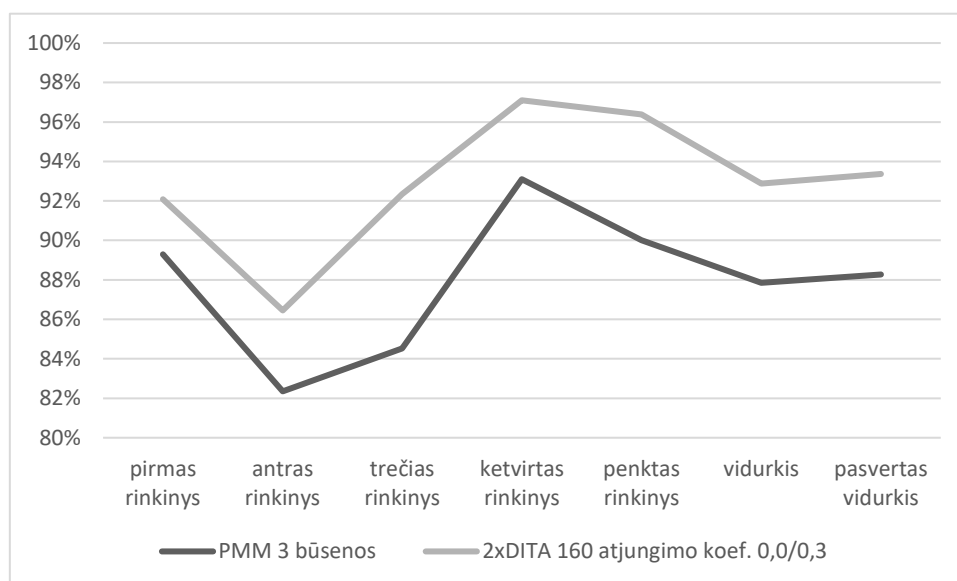


36 pav. Tiksliausių identifikavimo rezultatų palyginimas, naudojant paslėptus Markovo modelius



37 pav. Tiksliausių identifikavimo rezultatų palyginimas, naudojant rekurentinius neuroninius tinklus

Iš pateiktų lentelių sudarome tiksliausių identifikavimo rezultatų palyginamąjį grafiką tarp PMM ir RNT (38-as paveikslas).



38 pav. Pilnos eksperimento apimtys tiksliausių identifikavimo tikslumo rezultatų palyginimas

Kaip rodo pateiktas grafikas (38 pav.), lyginant du geriausius identifikavimo rezultatus leidusias pasiekti PMM ir RNT konfigūracijas ir topologijas, visų apmokymo ir testavimo imčių atveju diktorius tiksliau identifikuoja RT tipo klasifikatorius. Dar kartą pažymėsime, kad šios konfigūracijos ir topologijos buvo suboptimalios tarp to paties tipo modelių.

Rezultatų statistinis patikimumas. Siekiant nustatyti, ar gauti klasifikatoriaus identifikavimo tikslumo rezultatai yra statistiškai patikimi, apskaičiuojama patikimumo intervalo statistika, ji pateikiama 16-oje lentelėje. Šioje lentelėje yra pateikiamos kiekvieno klasifikatoriaus teisingų identifikavimo kiekių ir klaidų sumos. Pagal jas yra skaičiuojama McNemar testo statistika ir nustatoma p patikimumo intervalo reikšmė. Vertinant patikimumo intervalus, dažniausiai imama, kad patikimumas yra statistiškai reikšmingas, jei p reikšmė yra mažesnė už $\alpha=0,05$.

16 lentelė. Klaidų tipo detalizacija kiekvienam tyrimo rinkiniui ir jų statistinio patikimumo koeficientai

Parametro pavadinimas	Pirmas rinkinys	Antras rinkinys	Trečias rinkinys	Ketvirtas rinkinys	Penktas rinkinys
#1 – PMM ir RNT teisingi	1085	727	1273	1641	1022
#2 – PMM ir RNT neteisingi	38	52	61	24	31
#3 – tik PMM teisingas	66	118	99	33	77
#4 – tik RNT teisingas	100	129	190	100	91
McNemar testo statistika	6,5602	0,4048	28,027	32,751	1,0059
p -reikšmė 1 <i>tail</i>	0,00521	0,2623	0,0000	0,0000	0,1579
p -reikšmė 2 <i>tails</i>	0,0104	0,5245	0,0000	0,0000	0,3158

Išanalizavus pateiktą lentelę, matome, jog gauti patikimumo intervalai (p -reikšmė 2 *tails*) trimis atvejais yra mažesni už standartiškai naudojamą $\alpha=0,05$ koeficientą, todėl gauti rezultatai yra statistiškai reikšmingi trimis atvejais iš penkių.

Tyrimai naudojant kryžminį validavimą. Norint patikrinti, ar gautas tiksliausio klasikinio diktoriaus identifikavimo klasifikatoriaus ir RNT klasifikatoriaus rezultatas yra patikimas naudojant stratifikuoto atsitiktinio kryžminio validavimo metodą iš 10 pjūvių, eksperimentiniai tyrimai buvo pakartoti su pilnu LIEPA garsynu, padalintu į 5 dalis. Gauti rezultatai pateikiami 17-oje lentelėje.

17 lentelė. PMM ir RNT rezultatai naudojant kryžminio validavimo metodą

Klasifikatorius	Pirmas rinkinys	Antras rinkinys	Trečias rinkinys	Ketvirtas rinkinys	Penktas rinkinys
PMM 3 būsenos	92,27 proc. +/-0,94 proc.	85,75 proc. +/-1,32 proc.	89,41 proc. +/-0,64 proc.	96,37 proc. +/-0,72 proc.	93,41 proc. +/-1,31 proc.
2xDITA 160 atjungimo koef. 0,0/0,3	93,71 proc. +/-0,98 proc.	87,05 proc. +/-2,05 proc.	93,51 proc. +/-0,82 proc.	97,03 proc. +/-0,62 proc.	96,60 proc. +/-1,12 proc.

Kaip matome iš gautų rezultatų, neuroninio tinklo klasifikatorius visais atvejais diktorių identifikuoja tiksliau.

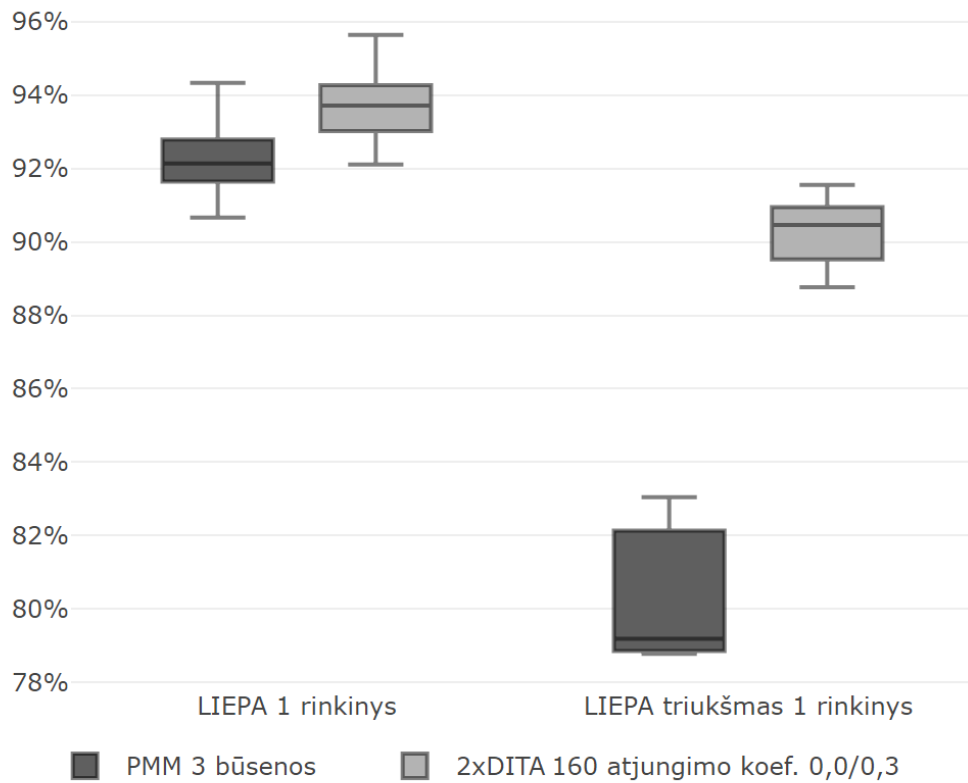
Tyrimai naudojant kryžminį validavimą su aptriukšmintais įrašais. Norint patikrinti, kaip pasiūlytas geriausias klasifikatoriaus variantas veiktų realiomis sąlygomis, tyrimai buvo atlikti naudojant modifikuotą aptriukšmintą LIEPA garsyną. Gauti diktoriaus identifikavimo tikslumo rezultatai pateikiami 18-oje lentelėje.

18 lentelė. PMM ir RNT rezultatai tiriant aptriukšmintą signalą kryžminio validavimo metodu

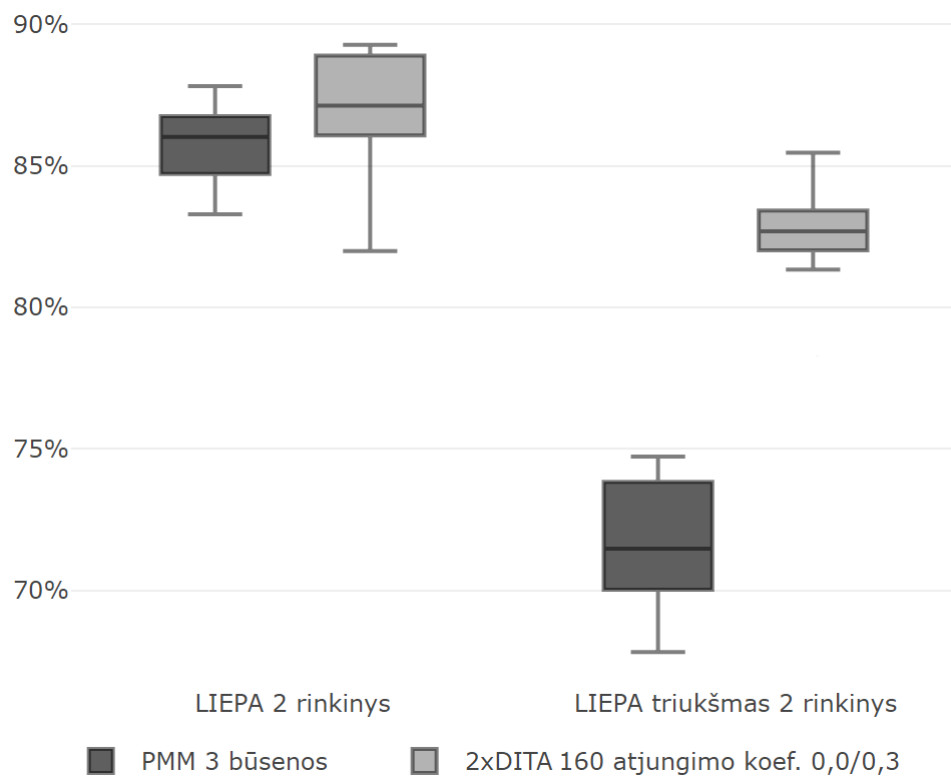
Klasifikatorius	Pirmas rinkinys	Antras rinkinys	Trečias rinkinys	Ketvirtas rinkinys	Penktas rinkinys
PMM 3 būsenos	80,06 proc. +/-1,63 proc.	71,61 proc. +/-2,41 proc.	77,35 proc. +/-1,52 proc.	82,50 proc. +/-1,47 proc.	71,95 proc. +/-2,01 proc.
2xDITA 160 atjungimo koef. 0,0/0,3	90,25 proc. +/-0,86 proc.	82,58 proc. +/-0,79 proc.	91,06 proc. +/-1,31 proc.	90,57 proc. +/-1,48 proc.	86,37 proc. +/-1,92 proc.

Pateiktuose rezultatuose yra stebimas didesnis neuroninio tinklo tikslumas nei klasikinio klasifikatoriaus atveju. Taip pat galime pastebėti, kad aptriukšminto signalo atveju tikslumo rezultato skirtumas tarp PMM ir DITA tinklų didėja, palyginus su švairiu įrašu.

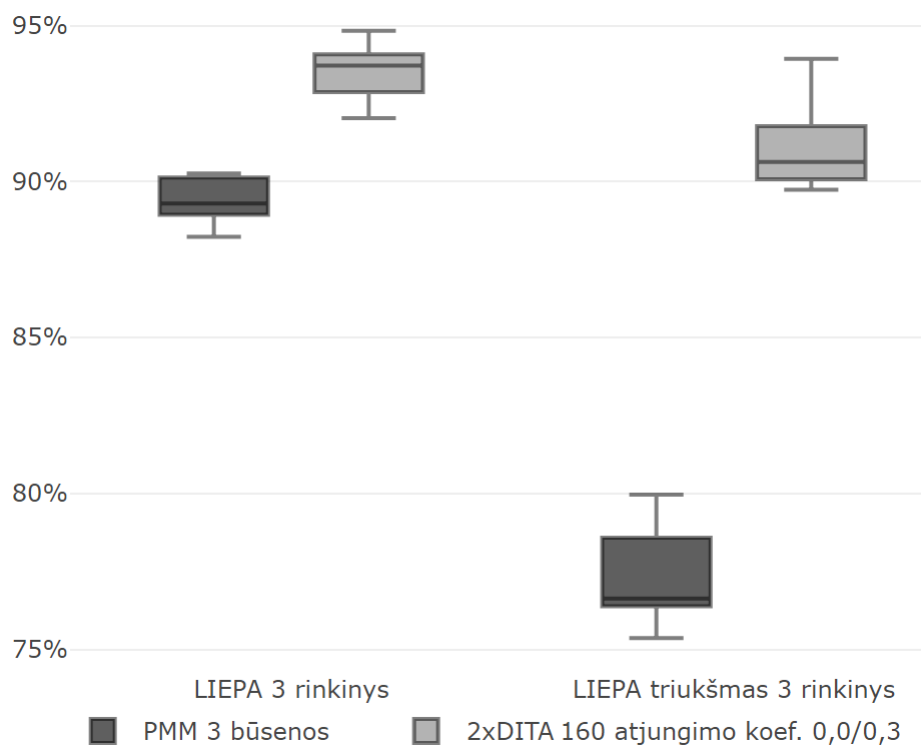
Kryžminio validavimo rezultatų vizualizacija. Palyginimui 38-43 paveiksluose pateikiami grafikai su kiekvieno diktorių rinkinio identifikavimo rezultatais naudojant kryžminio validavimo metodą. Grafikuose kairėje pusėje yra pateikiami nemodifikuoto diktorių rinkinio identifikavimo rezultatai pagal 17 lentelę, o dešinėje grafiko pusėje – identifikavimo rezultatai su to paties rinkinio aptriukšmintais signalais pagal 18 lentelę.



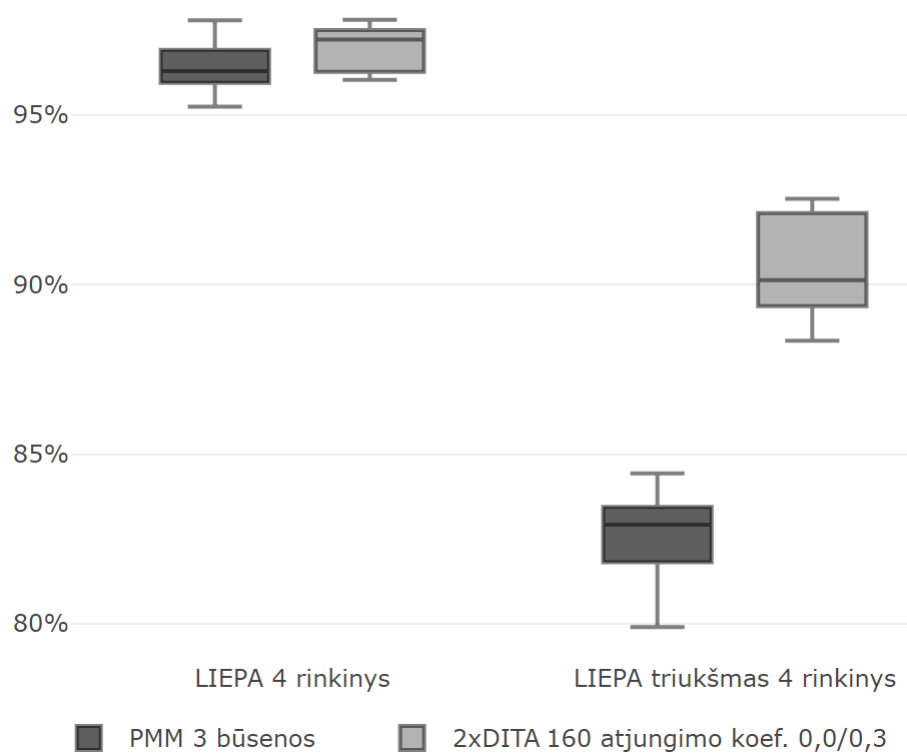
39 pav. Pirmo rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą



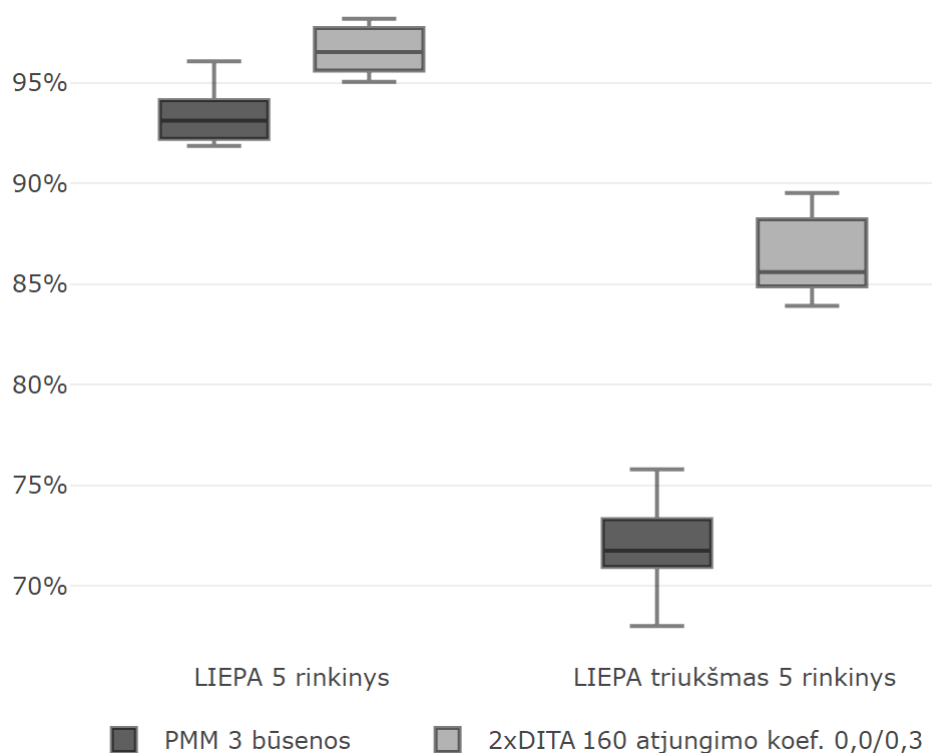
40 pav. Antro rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą



41 pav. Trečio rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą



42 pav. Ketvirto rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą



43 pav. Penkto rinkinio identifikavimo tikslumo rezultatų palyginimas naudojant kryžminį validavimą

Visais atvejais identifikavimo tikslumas yra didesnis naudojant siūlomą diktoriaus identifikavimo metodą, o iš pateiktų grafikų matyti, kad su aptriukšmintais rinkiniais RNT klasifikatoriaus pranašumas yra pastebimai didesnis.

3.3 Skyriaus išvados

Atlikus numatytus eksperimentus su LIEPA garsynu diktoriaus identifikavimo uždaviniui spręsti, nustatyta, jog:

- Atliekant diktoriaus identifikavimo tikslumo eksperimentinį tyrimą 5 skirtingiems diktorių rinkiniams, su pasiūlytu RNT klasifikatoriumi, tikslesnis ir statistiškai reikšmingas rezultatas gaunamas trimis atvejais iš penkių. Atlikus statistinio reikšmingumo testus, nustatyta, kad absoliutus identifikavimo tikslumas, naudojant RNT, yra geresnis visais tyrime atliktais atvejais, lyginant geriausius PMM ir RNT klasifikavimo metodus.
- Tikslesnis rezultatas yra stebimas tiek lyginant geriausius klasifikatorių variantus (nuo 3 proc. iki 6 proc.), tiek kryžminio validavimo metodą, tiek naudojant kryžminio validavimo metodą su aptriukšmintais įrašais.

Nustatyta, kad tikslesnis identifikavimas yra pasiekiamas naudojant RNT 2xDITA 160 atjungimo koef. 0,0/0,3 topologijos klasifikatorių.

IŠVADOS

1. Eksperimentais nustatyta, kad diktoriaus identifikavimo tikslumą galima padidinti naudojant ilgos trumpalaikės atminties rekurentinius neuroninius tinklus, palyginus su diktoriaus identifikavimui dažniausiai naudotais paslėptais Markovo modeliais. Diktoriaus identifikavimo tikslumo pagerinimas pasiektas naudojant tuos pačius požymių rinkinius, taigi vien dėl neuroninio tinklo klasifikavimo metodo naudojimo. Geriausias diktoriaus identifikavimo tikslumas naudojant paslėptuosius Markovo modelius svyruoja tarp 82,35 proc. ir 93,10 proc., tuo tarpu naudojant rekurentinius neuroninius tinklus svyruoja tarp 86,45 proc. ir 97,10 proc su tais pačiais duomenimis.
 - a. Pakartojus tyrimą naudojant atsitiktinio kryžminio validavimo metodą iš 10 pjūvių, gauti rezultatai su neuroniniu tinklu visais atvejais yra geresni nuo 0,66 iki 4,10 proc. nei naudojant paslėptuosius Markovo modelius.
 - b. Naudojant užtriukšmintus signalus, diktoriaus identifikavimo tikslumas naudojant rekurentinius Markovo modelius padidėja nuo 10,19 iki 14,42 proc., lyginant su paslėptais Markovo modeliais.
2. Tai leidžia teigti, kad pasiūlyta neuroninio tinklo klasifikatoriaus topologija, naudojant dvikryptės ilgos trumpalaikės atminties rekurentinį neuroninį tinklą lietuviškai kalbančių diktorių identifikavimui, yra pranašesnė negu paslėptais Markovo modeliais paremti diktoriaus identifikavimo metodai: naudojant rekurentinį neuroninį tinklą diktoriaus identifikavimo tikslumas padidėja nuo 3 proc. iki 6 proc., lyginant su tiksliausiu klasifikatoriumi, naudojančiu paslėptuosius Markovo modelius. Gauti rezultatai yra statistiškai reikšmingi ir statistinis reikšmingumas yra stebėtas trijose bandymų aibėse iš penkių. Naudojant kryžminio validavimo metodą, geresni rezultatai stebėti visų tyrimų atvejais.

3. Eksperimentiškai nustatyta, kad tarp įvairių konfigūracijų ilgos trumpalaikės atminties rekurentinių neuroninių tinklų (pavyzdžiui, tarp geriausių rezultatus parodžiusių dvikrypčio ilgos trumpalaikės atminties 160 neuronų tinklo ir dvikrypčio ilgos trumpalaikės atminties dviejų sluoksnių 160x160 tinklo topologijos), lietuviškai kalbančio diktoriaus identifikavimo tikslumas skiriasi tik 0,55 proc.

LITERATŪRA

Abdallah, S. J., Osman, I. M. ir Mustafa, M. E. 2012, „Text-Independent Speaker Identification Using Hidden Markov Model“, *World of Computer Science and Information Technology Journal (WCSIT)*, vol. 2, No. 6, p. 1-6.

Anderson, R. 2008, *Security Engineering*, Second edition. Wiley Publishing, Inc., Indianapolis, Indiana.

Arlot, S. ir Celisse, A. 2010, „A survey of cross-validation procedures for model selection“, *Statistics Surveys*, vol. 4, p. 40–79. Prieiga per internetą: <<https://arxiv.org/abs/0907.4728>>. [žiūrėta 2018 m. rugpjūčio 12 d.]

Aroon, A., ir Dhonde, S. B. 2015, „Speaker Recognition System using Gaussian Mixture Model“, *International Journal of Computer Applications*, vol. 130, No.14, p. 38-40.

Arsikere, H., Gupta, H. A. ir Alwan, A. 2014, „Speaker recognition via fusion of subglottal features and MFCCs“, *Interspeech-2014*. Prieiga per internetą: <https://www.isca-speech.org/archive/interspeech_2014/i14_1106.html>. [žiūrėta 2017 m. balandžio 30 d.]

Bawaskar, A. S. ir Kota, P. N. 2015, „Speaker Identification Based On MFCC and IMFCC Using GMM-UBM“, *IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)*, vol. 5, Issue 2, p. 53-60.

Bhattacharya, G., Alam, J., Stafylakis, T. ir Kenny, P. 2016, „Deep Neural Network based Text-Dependent Speaker Recognition: Preliminary Results“, *Odyssey 2016*, p. 9-15. Prieiga per internetą: <www.odyssey2016.org/papers/pdfs_stamped/67.pdf>. [žiūrėta 2017 m. rugsėjo 14 d.]

Bimbot, F., Bonastre, J. F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D. ir

De Boer, P. T., Kroese, D. P., Mannor, S. ir Rubinstein, R. Y. 2005, „A Tutorial on the Cross-Entropy Method“, *Annals of Operations Research*, vol. 134, no. 1, p. 19-67.

Castanedo, F. A. 2013, „Review of Data Fusion Techniques“, *The Scientific World Journal*, vol. 2013. Prieiga per internetą: <<http://dx.doi.org/10.1155/2013/704504>>. [žiūrėta 2018 m. gegužės 19 d.]

Chapelle, O., Scholkopf B. ir Zien, A. 2006, „Semi-Supervised Learning“, The MIT Press Cambridge, Massachusetts, London, England

Chauhan, N. 2017, „Speaker recognition using pattern recognition neural network and feedforward neural network“, *International Journal of Scientific & Engineering Research*, vol. 8, Issue 3, p. 1444-1446.

Chen, S. H. ir Luo, Y. R. 2009, „Speaker Verification Using MFCC and Support Vector Machine“, *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I*.

Cheng, S., Chen, I. F. ir Wang H. M. 2010, „Bayesian Speaker Recognition Using Gaussian Mixture Model and Laplace Approximation“, *Interspeech-2010*, p. 2730-2733. Prieiga per internetą: <www.isca-speech.org/archive/interspeech_2010/i10_2730.html>. [žiūrėta 2017 m. gegužės 11 d.].

Cortes, C. ir Vapnik, V. 1995, „Support Vector Networks“, *Machine Learning*, vol. 20, p. 273-297

Dietterich, T. G. 1998, „Approximate statistical tests for comparing supervised classification learning algorithms“, *Neural Computation*, vol. 10, p. 1895–1923. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/6790639/>>. [žiūrėta 2018 m. rugpjūčio 10 d.]

Deng, L. ir Yu, D. 2014, „Deep Learning: Methods and Applications“, *Foundations and Trends in Signal Processing*, vol. 7, no. 3-4, p. 197-387.

Deshmukh, S. D. ir Bachute, M. R. 2013, „Automatic Speech and Speaker Recognition by MFCC, HMM and Vector Quantization“, *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 3, Issue 1, p. 93-98.

Ding, I., Yen, C. T. ir Ou, D. C. 2014, „A Method to Integrate GMM, SVM and DTW for Speaker Recognition“, *International Journal of Engineering and Technology Innovation*, vol. 4, no. 1, 2014, p. 38-47.

Duda, R. O., Hart, P. E. ir Stork, D. G. 2000, *Pattern Classification*, Second Edition, John Wiley & Sons, New York.

Fakotakis, N., Georgila, K. ir Tsopanoglou, A., A. 1997, „A Continuous HMM text-independent speaker recognition system based on vowel spotting“, *EUROSPEECH'97, 5th European Conference on Speech Communication and Technology*. Prieiga per internetą: <https://www.isca-speech.org/archive/eurospeech_1997/e97_2347.html>. [žiūrėta 2017 m. rugsėjo 4 d.].

Fawcett, T. 2005, „An introduction to ROC analysis“, *Pattern Recognition Letters*, vol. 27, no. 8, p. 861-874.

Gal, Y. ir Ghahramani, Z. 2016, *A Theoretically Grounded Application of Dropout in Recurrent Neural Networks*, University of Cambridge. Prieiga per internetą: <<https://arxiv.org/abs/1512.05287?context=stat>>. [žiūrėta 2017 m. gegužės 1 d.].

Ganjeizadeh, F., Lei, H., Maganito, A. ir Pallipatta, G. 2014, „Reducing the Computational Complexity of the GMM-UBM Speaker Recognition Approach“, *International Journal of Engineering Research & Technology*, vol. 3, no. 3, p. 1793-1797.

Garcia-Romero. D., Zhang, X., Alan McCree, A. ir Povey, D. 2014, „Improving speaker recognition performance in the domain adaptation challenge using deep

neural networks“ in *Spoken Language Technology Workshop (SLT), South Lake Tahoe, Nevada, USA, December 7-10, 2014*, South Lake Tahoe, Nevada (NV).

Ge, Z., Iyer, A. N., Cheluvvaraja, S., Sundaram, R. ir Ganapathiraju, A. 2017, „Neural Network Based Speaker Classification and Verification Systems with Enhanced Features“, *Intelligent Systems Conference 2017*. Prieiga per internetą: <<https://arxiv.org/abs/1702.02289>>. [žiūrėta 2017 m. rugsėjo 5 d.]

Glorot, X. ir Bengio, Y. 2010, „Understanding the difficulty of training deep feedforward neural networks“, in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Chia Laguna Resort, Sardinia, Italy, p. 249-256.

Goodfellow, I., Bengio, Y. ir Courville A. 2016, *Deep Learning*, The MIT Press, Cambridge.

Graves, A., Mohamed, A. ir Hinton, J. 2013, „Speech recognition with deep recurrent neural networks“ in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, May 26-31, 2013*, Vancouver, BC, Canada.

Graves, A. ir Schmidhuber, J. 2005, „Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures“ in *2005 IEEE International Joint Conference on Neural Networks (IJCNN), Montreal, Canada, July 31-August 4, 2005*, Montreal, Que., Canada.

Graves, A. 2012, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, Berlin.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R. ir Schmidhuber, J. 2015, „LSTM: A Search Space Odyssey“, *Cornell University Library*. Prieiga per internetą: <<https://arxiv.org/abs/1503.04069v1>>. [žiūrėta 2017 m. rugsėjo 14 d.].

Hasan, M. H., Jaafar, H. ir Ramli, D. A. 2012, „Evaluation on Score Reliability for Biometric Speaker Authentication Systems“, *Journal of Computer Science*, vol. 8, no. 9, p. 1554-1563. Prieiga per internetą: <<http://thescipub.com/abstract/10.3844/jcssp.2012.1554.1563>>. [žiūrėta 2018 m. gegužės 19 d.].

Himawan, I., Motlicek, P., Imseng, D., Potard, B., Kim, N., ir Lee, J. 2015, „Learning feature mapping using deep neural network bottleneck features for distant large vocabulary speech recognition“, *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4540-4544. Prieiga per internetą: <<http://publications.idiap.ch/index.php/publications/show/3117>>. [žiūrėta 2016 m. rugsėjo 7 d.].

Hochreiter, S. ir Schmidhuber, J. 1997, „Long short-term memory“, *Neural Computation*, vol. 9, no 8, p. 1735-1780.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V. ir Woodland, P. 2005, *The HTK Book*, Cambridge University Engineering Department, Cambridge.

Islam, M., Khan, F. H. ir Haque, A. A. M. 2013, „A Novel Approach for Text-Independent Speaker Identification Using Artificial Neural Network“, *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 4, p. 838-844.

Jayanth, M. ir Roja, R. B. 2016, „Speaker Identification based on GFCC using GMM-UBM“, *International Journal of Engineering Science Invention*, vol. 5, no. 5, p. 62-65.

Jain A., Dass C. ir Nandakumar K. 2004, „Can soft biometric traits assist user recognition?“, *Proceedings of SPIE*, vol. 5404, p. 561-572.

Jain A., Ross A. ir Prabhakar S. 2004, „An Introduction to Biometric Recognition“, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, p. 4-20.

Kamarauskas, J. 2009, *Asmens atpažinimas pagal balsą. Doktoro disertacija*, Technika, Vilnius

Kanagasundaram, A., Vogt, R., Dean, D., Sridharan, S. ir Mason, M. 2011, „i-vector Based Speaker Recognition on Short Utterances“, *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, p. 2341-2344. Prieiga per internetą: <<http://eprints.qut.edu.au/46313/>> [žiūrėta 2015 m. lapkričio 17 d.]

Kenny, P. 2006, „Joint Factor Analysis of Speaker and Session Variability: Theory and Algorithms“, *Technical report, Computer Research Institute of Montreal*. Prieiga per internetą: <<http://www.crim.ca/perso/patrick.kenny/FAtheory.pdf>>. [žiūrėta 2016 m. spalio 8 d.].

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M. ir Tang, P. T. P. 2017, „On large-batch training for deep learning: generalization gap and sharp minima“, *International Conference on Learning Representations ICLR 2017*. Prieiga per internetą: <<https://arxiv.org/abs/1609.04836>>. [žiūrėta 2017 m. balandžio 25 d.].

Kingma, D. P. ir Ba, J. L. 2015, „Adam: A method for stochastic optimization“, in *3rd International Conference for Learning Representations, San Diego, USA, July 23, 2015*, San Diego.

Kinnunen T. ir Li H. 2009, „An Overview of Text-Independent Speaker Recognition: from Features to Supervectors“, *Speech Communication*, vol. 52, no. 1, p. 12-40. Prieiga per internetą: <<https://doi.org/10.1016/j.specom.2009.08.009>>. [žiūrėta 2017 m. spalio 5 d.].

Krose, B. ir Smagt, P. 1996, *An Introduction to Neural Networks*, Eight edition, The University of Amsterdam, Amsterdam.

Kumar, K., Kim, C. ir Stern, R. M. 2011, „Delta-spectral cepstral coefficients for robust speech recognition“ in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011, Prague, Czech Republic, May 22-27, 2011*, Prague Congress Center, Prague.

Laurinčiukaitė, S., Telksnys, L., Kasparaitis, P., Kliukienė, R. ir Paukštytė, V. 2017, *Lithuanian Speech Corpus Liepa for the Development of Lithuanian Speech Controlled Equipment*. Pateiktas spaudai.

Lei, Y., Scheffer, N., Ferrer, L. ir McLaren, M. 2014, „A novel scheme for speaker recognition using a phonetically-aware deep neural network“ in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, May 4-9, 2014*, Florence, Italy.

Lipeika, A. L. 2005, „Formantiniai požymiai atpažįstant kalbą“, *Informacijos mokslai*, nr. 34.

Lv, G. ir Zhao, H. 2012, „Joint Factor Analysis of Channel Mismatch in Whispering Speaker Verification“, *Archives of acoustics*, vol. 37, no. 4, p. 555–559.

Mahola, U., Nelwamondo F. V. ir Marwala, T. 2007, „HMM Speaker Identification Using Linear and Non-linear Merging Techniques“, *Cornell University Library*. Prieiga per internetą: <<https://arxiv.org/abs/0705.1585>>. [žiūrėta 2017 m. rugsėjo 4 d.].

Meglouli H. ir Khebli A. 2015, *Speaker recognition by gaussian mixture model (GMM)*. Prieiga per internetą: <<http://fhc.univ-boumerdes.dz/index.php/en/>>. [žiūrėta 2017 m. rugsėjo 4 d.].

Mohanty, S. ir Swain, B. K. 2014, „Speaker Identification from Oriya Voices Using SVM“, *International Journal of Innovations & Advancement in Computer Science IJIACS*, vol. 3, no. 8, p. 33-39.

Molau, S., Pitz, M., Schluter, R. ir Ney, H. 2001, „Computing mel-frequency cepstral coefficients on the power spectrum“ in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP '01)*, Salt Lake City, USA, May 7-11, 2001, Salt Lake City, UT, USA.

Moreno, P. J. ir Ho, P. P. 2003, „A New SVM Approach to Speaker Identification and Verification Using Probabilistic Distance Kernels“ in *Eurospeech 2003, Geneva, Switzerland, September 1-4, 2003*, Geneva, Switzerland.

Muda, L., Begam, M. ir Elamvazuthi, I. 2010, „Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques“, *Journal of Computing*, vol. 2, no. 3, p. 138-143. Prieiga per internetą: <<https://sites.google.com/site/journalofcomputing/>>. [žiūrėta 2018 m. gegužės 12 d.].

Naik, J. 1990, „Speaker Verification: A Tutorial“, *IEEE Communication Magazine*, vol. 28, no. 1, p. 42-48.

Narayana, L. ir Kopparapu, S. K. 2014, „Choice of Mel Filter Bank in Computing MFCC of a Resampled Speech“. Prieiga per internetą: <<https://arxiv.org/abs/1410.6903>>. [žiūrėta 2018 m. rugpjūčio 10 d.]

Narayanaswamy, B. 2005, „Improved Text-Independent Speaker Recognition using Gaussian Mixture Probabilities“, *A Report in Candidacy for the Degree of Master of Science (M.S.)*, Department of Electrical and Computer Engineering Carnegie Mellon University, Pittsburg, Pennsylvania.

Nath, D. ir Kalita, S. 2014, „Feature Selection Method for Speaker Recognition using Neural Network“, *International Journal of Computer Applications*, vol. 101, no. 3, p. 38-44.

Nijhawan, G. ir Soni, M. K. 2014, „Speaker Recognition using Support Vector Machine“, *International Journal of Computer Applications*, vol. 87, no. 2, p. 7-10.

Omar, M. K. ir Pelecanos, J. 2010, „Training Universal Background Models for Speaker Recognition“, *The Speaker and Language Recognition Workshop Odyssey-2010*, p. 52-57. Prieiga per internetą: <https://www.isca-speech.org/archive_open/odyssey_2010/od10_010.html>. [žiūrėta 2015 m. lapkričio 17 d.].

Picone, W. J. 1993, „Signal Modeling Techniques in Speech Recognition“, *Proceedings of the IEEE*, vol. 81, no. 9, p. 1215-1247. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/237532/>>. [žiūrėta 2018 m. gegužės 12 d.].

Pham, V., Bluche, T., Kermorvant, C. ir Louradour, J., 2014, „Dropout Improves Recurrent Neural Networks for Handwriting Recognition“, in *2014 14th International Conference on Frontiers in Handwriting Recognition*, Heraklion

Rabiner, L. R. 1989, „A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition“, *Proceedings of the IEEE*, vol. 77, no. 2, p. 257-86. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/18626/>>. [žiūrėta 2018 m. rugpjūčio 10 d.]

Ravindran, S., Anderson, V. D. ir Slaney, M. 2006, „Improving the Noise-Robustness of Mel-Frequency Cepstral Coefficients for Speech Processing“ in *ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition SAPA2006, Pittsburg, USA, September 16, 2006*, Pittsburg PA, USA.

Reynolds, D. A., Quatieri, F. T. ir Dunn, R. B. 2000, „Speaker Verification Using Adapted Gaussian Mixture Models“, *Digital Signal Processing*, vol. 10, p. 19-41.

Reynolds, D. A. ir Rose, R. 1995, „Robust text independent speaker identification using Gaussian mixture speaker models“, *IEEE transactions on speech and audio processing*, vol. 3, no. 1, p. 72-83.

Reynolds, D. A. 1995, „Speaker identification and verification using Gaussian mixture models“, *Speech communication*, vol. 17, p. 91-108.

Reynolds, D. A. 2004, „A Tutorial on Text-Independent Speaker Verification“, *EURASIP Journal on Applied Signal Processing*, vol. 4, p. 430-451. Prieiga per internetą: <<https://doi.org/10.1155/S1110865704310024>>. [žiūrėta 2018 m. gegužės 29 d.]

Richardson, F., Reynolds, D. ir Dehak, N. 2015, „Deep Neural Network Approaches to Speaker and Language Recognition“, *IEEE signal processing letters*, vol. 22, no. 10, p. 1671-1675.

Ringelienė, Ž. ir Filipovič, M. 2011, „Žodžių atpažinimo, grįsto paslėptaisiais Markovo modeliais, vizualizavimo ir analizės programinė įranga“, *Informacijos mokslai*, no. 56, p. 63-72.

Robinson, S. 2011, „Using Biometry for Security and Identification“, *interaktyvus*. Prieiga per internetą: <<http://www.brighthouse.com/computing/smb-security/articles/63325.aspx>>. [žiūrėta 2015 m. vasario 22 d.].

Rubal ir Mehan, V. 2017, „Isolated Word Recognition and Sign Language Detection using LPC and MFCC“, *International Journal of Advance Research in Computer Science and Management Studies*, vol. 5, no. 7, p. 137-145.

Russell, S. J. ir Norvig, P. 2009, *Artificial Intelligence: A Modern Approach*, Third Edition., Prentice Hall, New Jersey.

Saxe, A. M., McClelland, J. L. ir Ganguli, S. 2014, „Exact solutions to the nonlinear dynamics of learning in deep linear neural networks“, *Cornell University Library*. Prieiga per internetą: <<https://arxiv.org/abs/1312.6120>>. [žiūrėta 2018 m. rugpjūčio 10 d.].

Schmidhuber, J. 2014, „Deep Learning in Neural Networks: An Overview“, *Neural Networks*, vol. 61, p. 85-117. Prieiga per internetą: <<https://arxiv.org/abs/1404.7828>>. [žiūrėta 2018 m. gegužės 28 d.].

Schuster, M. ir Paliwal, K. K. 1997, „Bidirectional Recurrent Neural Networks“, *IEEE Transactions on Signal Processing*, vol. 45, no. 11, p. 2673-2681. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/650093/>>. [žiūrėta 2018 m. gegužės 28 d.].

Senoussaoui, M., Kenny, P., Dehak, N. Ir Dumouchel, P. 2010, „An i-vector Extractor Suitable for Speaker Recognition with both Microphone and Telephone Speech“ in *Proc. Odyssey Speaker and Language Recognition Workshop, 2010, Brno, Check Republic, June 28, 2010*, Brno, Check Republic.

Shavlik, J. W., Dietterich, T. ir Dietterich, T. G. 1990, *Readings in machine learning*. Kaufmann Publishers, California.

Srinivas, V., Rani, S. ir Madhu, T. 2014, „Neural network based Classification for Speaker Identification“, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol.7, no.1, p. 109-120.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. ir Salakhutdinov, R. 2014, „Dropout: A Simple Way to Prevent Neural Networks from Overfitting“, *Journal of Machine Learning Research*, vol. 15, p. 1929-1958.

Smith, S. W. 2003, *Digital Signal Processing: A Practical Guide for Engineers and Scientists*, 1st Edition, Elsevier Inc, Burlington.

Šalna B. ir Kamarauskas J. 2010, „Evaluation of Effectiveness of Different Methods in Speaker Recognition“, *Elektronika ir elektrotechnika*, vol. 2, no. 98, p. 67-70.

Tiwari, V. 2010, „MFCC and its applications in speaker recognition“, *International Journal on Emerging Technologies*, vol. 1, no. 1, p. 19-22.

Togneri, R. ir Pullella, D. 2011, „An Overview of Speaker Identification: Accuracy and Robustness Issues“, *IEEE Circuits and Systems Magazine*, vol. 11, no. 2, p. 23-61. Prieiga per internetą: <<https://ieeexplore.ieee.org/document/5871484/>>. [žiūrėta 2018 m. gegužės 18 d.].

Variani, E., Lei, X., McDermott, E., Lopez Moreno, I. ir Gonzalez-Dominguez, J. 2014, „Deep Neural Networks for Small Footprint Text-dependent Speaker Verification“ in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Florence, Italy, May 4-9, 2014 : Proceedings*, Florence, Italy.

Zazo, R., Lozano-Diez, A., Gonzalez-Dominguez, J., Toledano, D. ir Gonzalez-Rodriguez, J. 2016, „Language Identification in Short Utterances Using Long Short-Term Memory (LSTM) Recurrent Neural Networks“, *PloS ONE*, vol. 11, no. 1, p. 1-17. Prieiga per internetą: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146917>>. [žiūrėta 2017 m. rugsėjo 16 d.].

Zheng, F., Zhang, G. ir Song, Z. 2001, „Comparison of Different Implementations of MFCC“, *Journal of Computer Science and Technology*, vol. 16, no. 6, p. 582-589. Prieiga per internetą: <https://www.researchgate.net/publication/220584952_Comparison_of_Different_Implementations_of_MFCC>. [žiūrėta 2018 m. gegužės 12 d.].

Zheng, R., Ulag, S. ir Xu, B. 2004, „Text-independent speaker identification using GMM-UBM and frame level likelihood normalization“ in *2004 International Symposium on Chinese Spoken Language Processing, Hong Kong, China, December 15-18, 2004*, Hong Kong, China.

AUTORIAUS PUBLIKACIJŲ SĄRAŠAS DISERTACIJOS TEMA

Straipsniai publikuoti periodiniuose recenzuojamuose leidiniuose:

1. Dovydaitis, L. ir Rudžionis, V. 2018, „Building bi-directional LSTM neural network based speaker identification system“, *Computational Science and Techniques*, vol. 6, no 1, p. 574–580. DOI: 10.15181/csat.v6i1.1579
2. Dovydaitis, L. ir Rudžionis, V. 2018, „Speaker identification accuracy improvement using BLSTM neural network“, *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 9, no. 2, p. 31–37. DOI: 10.21817/indjcse/2018/v9i2/180902017

Straipsniai recenzuojamuose konferencijų leidiniuose:

1. Dovydaitis L. ir Rudžionis V. 2017, „Identifying Lithuanian Native Speakers Using Voice Recognition“ in Abramowicz W. (eds) *Business Information Systems Workshops BIS 2017. Lecture Notes in Business Information Processing*, p. 79–84. DOI: 10.1007/978-3-319-69023-0_8
2. Dovydaitis L., Rasytas T. ir Rudžionis V. 2017, „Speaker Authentication System Based on Voice Biometrics and Speech Recognition“ in Abramowicz W., Alt R., Franczyk B. (eds) *Business Information Systems Workshops BIS 2016. Lecture Notes in Business Information Processing*, p. 79–84. DOI: 10.1007/978-3-319-52464-1_8
3. Dovydaitis, L. ir Rudžionis, V. 2015, „Asmens balso panaudojimas autentikavimui“ in *Informacinės technologijos 2015, 20-oji tarpuniversitetinė magistrantų ir doktorantų konferencija, Kaunas, Lietuva, balandžio 24 d., 2015 : konferencijos pranešimų medžiaga*, Kaunas, Lietuva, p. 147–151.

Santraukos konferencijų leidiniuose:

1. Dovydaitis, L. ir Rudžionis, V. 2017, „Speaker Recognition Using Deep Neural Networks“ in *Data analysis methods for software systems: 9th*

- International Workshop, Druskininkai, Lithuania, November 30 – December 2, 2017, Druskininkai, Lithuania, p. 15.*
2. Dovydaitis, L. ir Rudžionis, V. 2016, „Speaker Recognition Using Deep Neural Networks“ in *Data analysis methods for software systems: 8th International Workshop, Druskininkai, Lithuania, December 1–3, 2016, Druskininkai, Lithuania, p. 18.*
 3. Dovydaitis, L. ir Rudžionis, V. 2015, „GMM-UBM Enhancement with DNNs for Automated Speaker Recognition Systems“ in *Data analysis methods for software systems: 7th International Workshop, Druskininkai, Lithuania, December 3–5, 2015, Druskininkai, Lithuania, p. 18.*

Laurynas Dovydaitis

LIETUVIŠKAI KALBANČIO DIKTORIAUS IDENTIFIKAVIMO
NAUDOJANT GRĮŽTAMOJO RYŠIO NEURONINIUS TINKLUS
TIKSLUMO TYRIMAS

Daktaro disertacijos santrauka

Technologijos mokslai

Informatikos inžinerija (07 T)

Redaktorė Dagnė Pakalniškytė-Alseikienė