



## TURINYS

<b>Žymenys ir santrumpos</b>	<b>8</b>
<b>1. Įvadas. Užduoties analizė</b>	<b>9</b>
<b>2. Informacinių Parkinsono ligos diagnostikos sistemų apžvalga</b>	<b>12</b>
2.1. Tiesinių jėgos jutiklių metodas	12
2.2. Dirbtinių neuronų tinklų metodas	13
2.3. Atraminių vektorių mašinos metodas	14
2.4. Statistinės analizės metodas	16
2.5. Parkinsono ligos stadijos nustatymo metodas	17
<b>3. Signalų analizės programos kūrimas</b>	<b>20</b>
3.1. Bendros programos struktūrinės schemos sudarymas	20
3.2. Pirminio signalų apdorojimo programos kūrimas	21
3.3. Požymių išskyrimo programos kūrimas	24
3.3.1. Matmenų mažinimas	24
3.3.2. Požymių tyrimas	26
3.4. Požymių klasifikavimo programos kūrimas	33
3.5. Duomenų analizės programos kūrimas	37
<b>4. Signalų analizės programos įgyvendinimas</b>	<b>39</b>
4.1. Bendro programos algoritmo schemos sudarymas	39
4.2. Požymių klasifikavimo programos algoritmo schemos sudarymas	44
4.3. Signalų analizės programos įgyvendinimas	46
<b>5. Signalų analizės programos patikra</b>	<b>47</b>
5.1. Eksperimentų plano rengimas	47
5.2. Duomenų eksperimentams rengimas	48
5.3. Programos patikros rezultatai	48
<b>6. Rezultatų apibendrinimas</b>	<b>52</b>
<b>Literatūros ir informacinių šaltinių sąrašas</b>	<b>53</b>
<b>Santrauka</b>	<b>56</b>
<b>PRIEDAI</b>	<b>57</b>



## Žymenys ir santrumpos

- VŽRJ – Vertikali žemės reakcijos jėga (angl. *Vertical Ground Reaction Force*);
- PCA – principinė komponentų analizė (angl. *Principal Component Analysis*);
- LDA – tiesinė diskriminanto analizė (angl. *Linear Diskriminant Analysis*);
- SVM – vektoriaus palaikymo mašina (angl. *Support Vector Machine*);
- STC – signalo nuokrypio skaitiklis (angl. *Signal Turn Counter*);
- Co – kontrolinis subjektas, kurio ligos istorijoje nebuvo užregistruota jokių neurologinių susirgimų;
- Pt – Parkinsono liga sergantis subjektas;
- Hz – ciklų skaičius per vieną sekundę;
- s – sekundė;



## 1. Įvadas. Užduoties analizė

Bakalauro baigiamojo darbo tema – Parkinsono ligos eigos stebėjimo programa. Parkinsono liga yra dažniausiai pasitaikantis neurodegeneracinis judėjimo sutrikimas. Ankstyva ligos diagnozė ir efektyvus terapijos stebėjimas yra būtinas pacientų gydymui ir ligos. Šiuo metu neegzistuoja gydytojų patvirtintos objektyvios ir vieningos vertinimo sistemos, kuri tiksliai atpažintų Parkinsono ligos simptomus. Vienas iš didžiausiai pasireiškiančių simptomų yra eisenos sutrikimas. Sutrikimo dažnumą ir svarbą patvirtina viešai prieinama duomenų bazė, kurioje yra pateikiami sveikų ir sergančių Parkinsono liga žmonių eisenos duomenys. Duomenų bazė vadinasi “PhysioBank” [14].

Darbo tikslas – sukurti programą, kuri atpažintų sveiką ir Parkinsono liga sergantį žmogų, remiantis vertikalios jėgos jutikliais gautais signalais. Programa įgyvendinta Matlab platformoje. Ji pasirinkta dėl plataus įrankių kiekio, kuris yra įgyvendintas Matlab aplinkoje. Nurodytoje platformoje taip pat yra labai patogų ir greitą realizuoti signalų apdorojimo sistemas dėl jos architektūros – visi kintamieji yra matricos.

Šiuo metu Parkinsono liga yra aktyviai tiriama [1], kadangi ją susirgęs žmogus praranda galimybę laisvai gyventi. Motorinės funkcijos sutrikimai padaro gyvenimą problematišką: drebulys sukelia nepatogumų atliekant didesnio kruopštumo reikalaujančius darbus, galūnių sustingimas sukelia didelių nepatogumų atliekant paprasčiausius kasdienes procedūras. Toliau pasireiškia psichiniai sutrikimai, dėl neurologinio sutrikimo: depresija, apatija, miego sutrikimas. Tik ankstyva ligos diagnozė gali sulėtinti ligos plitimą ir pagelbėti ją susirgusiems žmonėms. Dėl šios priežasties darbo tema pasirinkta Parkinsono liga.

Darbe naudojami signalai, kuriuos pateikia jėgos jutikliai. Jie matuoja vertikalią žemės reakcijos jėgą (VŽRJ) (angl. *Vertical Ground Reaction Force – vGRF*). Fizikoje, ir būtent biomechanikoje, VŽRJ nurodo kokia jėga žemė atsako ją veikiančiam objektui. Kaip pavyzdžiui, stovinti žmogus slegia žemę jėga, kuri lygi jo masei ir tuo pat metu, žemė slegia žmogų priešinga, lygiai tokia pačia jėga. Jutikliai parinkti dėl šaltinyje [20] pateiktos informacijos, kad VŽRJ yra geras žmogaus kūno stabilumo indikatorius. Duomenų surinkimo platformos ir duomenų surinkimo sesijų atlikinėti nereikėjo, kadangi naudojami duomenys yra prieinami duomenų bazėje.

Didžiausia ligos atpažinimo problema slypi signalo požymių, kurios geriausiai identifikuos sergantį Parkinsono liga nuo sveiko, nesergančio asmens, parinkime. Darbe apžvelgtos keli galimi signalo požymiai, kurių reikšmės geriausiai identifikuos skirtumus tarp signalų, remiantis analoginių sistemų apžvalga. Iš galimų požymių sudaroma požymių erdvė, kurią naudojantis atliekamas subjektų atpažinimas. Darbe panaudotas Matlab aplinkoje įgyvendintas įrankis, kuris skirtas suprojektuoti naują matmenų plokštumą, kuri įgyvendino Wojtech Franc savo magistriniam darbe [13].

Sukurtas produktas pateikia diagnozės rezultatą – ar subjektas turi Parkinsono liga ser-

gančių subjektų eisenos požymių ar jų neturi. Produktas neatsižvelgs į kitus ligos simptomus: drebulys (rankų, kojų, žandikaulio, galvos), standumas (galūnių arba liemens sustingimas), bradikinezija (judesių lėtumas), pozicijos nestabilumas (arba sutrikęs balansas). Pati programa duomenis analizuos jau po duomenų surinkimo. Tai reiškia, kad pirmiausiai duomenys yra surenkami, o vėliau įkeliami į programą tolimesniam apdorojimui.

Parkinsono ligos eigos stebėjimo programa – sukurtas algoritmas, įgyvendintas kompiuterine programa. Pačiam kompiuteryje turi būti veikiantis Matlab programinis paketas. Programa parašyta naudojantis Matlab 7.12.0 (R2011) versija su “Neural Network Toolbox” ir “Statistical Toolbox” įrankiu. Eigos stebėjimu turima omenyje visuomet egzistuojantį neapibrėžtų, galimą programos netikslumą. Visiškai programa remtis, diagnozuojant Parkinsono ligą, nėra galima, kadangi, kaip jau minėta anksčiau – eigos sutrikimas nėra vienintelis ligos požymis. Turi būti atlikti ir kiti tyrimai, norint tiksliai diagnozuoti ligą.

Darbo objekto sudėtis yra asmeniniam kompiuteriui skirta programa, jėgos jutiklių signalų generavimo programa. Kokiam kompiuteriui kuriama programa, paminėta darbo užduoties aprašyme. Jėgos jutiklių generavimo programa (modulis) atsakinga už signalų nuskaitymą iš duomenų bazės ir jų pateikimą sistemos algoritmui. Taip pat signalų generavimo programa (modulis) naudojamas programos demonstracinei versijai įgyvendinti.

Duomenys sistemai pateikiami iš “PhysioBank” duomenų bazės [14]. Duomenys duomenų bazėje surinkti diskretizuojant jėgos jutiklių signalus 100 Hz diskretizavimo dažniu. Kiekvienu laiko momentu yra įrašoma nauja eilutė į duomenų tekstinę bylą. Eilutę sudaro 19 skilčių:

- Skiltis 1 nurodo laiką (sekundėmis);
- Skiltys 2–9 nurodo kairės kojos 8 jutiklių VŽRJ, Niutonais;
- Skiltys 10–17 nurodo dešinės kojos 8 jutiklių VŽRJ, Niutonais;
- Skiltis 18 nurodo kairės kojos suminę VŽRJ, Niutonais;
- Skiltis 19 nurodo dešinės kojos suminę VŽRJ, Niutonais;

Duomenys surinkti: “Galit Yogev et al” (dual tasking in PD; Eur J Neuro, 2005), “Hausdoff et al” (RAS in PD; Eur J Neuro, 2007), “Silvi Frenkel-Toledo et al” (Treadmill walking in PD; Mov Disorders, 2005). Duomenų rinkimo sesijų autorių vardais pažymimas pirmosios bylos pavadinimo dvi raidės: “Ga” priklauso “Galit Yogev et al”, “Ju” – “Hausdoff et al”, “Si” – “Silvi Frenkel-Toledo et al”. Sesijose dalyvavo du subjektų tipai: kontroliniai (sveiki) subjektai ir Parkinsono liga sergantys subjektai. Bylojos pavadinime jie pažymėti “Co” ir “Pt” raidžių rinkiniais. Sekantis pirmas numeris nurodo subjekto atpažinimo numerį, po brūkšnio einantis antras numeris nurodo subjekto duomenų rinkimo seką. Aprašyme nurodyta, kad visuose duomenų rinkimų sesijose iš viso dalyvavo 93 sergantis subjektas ir 73 Parkinsono liga sergantis subjektas.

Programos veikimas, atliekamas klasifikavimo mechanizmu, vertinamas taiklumu ir jautru-



mu. Parametrai yra apskaičiuojami iš pasiklovimo matricos.

## 2. Informacinių Parkinsono ligos diagnostikos sistemų apžvalga

Šiame skyriuje apžvelgtos analoginės informacinės sistemos arba bandymai sukurti sistemą, kuri, remiantis įvairių jutiklių pagalba, gebėtų atpažinti Parkinsono ligą. Apibendrintai aptarti nagrinėjamų sistemų teigiamos ir neigiamos savybės – ar pasirinkti požymiai yra argumentuoti ar naudojamas klasifikatorius yra parinktas atsižvelgus į naudojamą požymių erdvę, kokie gauti rezultatai.

## 2.1. Tiesinių jėgos jutiklių metodas

Pirmas apžvalgai darbas yra “Gait feature extraction in Parkinson’s disease using low-cost accelerometers” [19], nėra naudojama vertikali žemės reakcijos jėga, skaičiuojant požymius. Sprendime panaudoti pigūs, keturi tiesinio pagreičio jutikliai. Toks sprendimas yra lengvesnis, už prieš tai apžvelgtą, kadangi jame panaudoti šešiolika vertikalios žemės reakcijos jutiklių. Nagrinėjamos Parkinsono ligos eisenos stingimo ir nesimetriškumo požymiai. Iš pirminės sprendimo apžvalgos iškarto išryškėja sprendimo trūkumas – tiriant galimus požymius, panaudotas tik vienas Parkinsono liga sergantis ir vienas kontrolinis subjektas. Taip atliekant tyrimą, egzistuoja didelė tikimybė, kad visi gaunami signalo eisenos požymiai gali būti būdingi tik nagrinėjamam subjektui, tačiau gali visiškai negalioti kitiems subjektams ar subjektų grupei. Egzistuoja ir tokia tikimybė, kad ištirti požymiai, kurie galioja Parkinsono subjektui, gali galioti ir tyrime nedalyvavusiam kontroliniam subjektui, kadangi panaudotas tik vienas kontrolinis subjektas tyrimo metu.

Nepriklausomai nuo atlikto tyrimo siaurumo, darbe pateiktos svarbios eisenos analizės išvados. Kontrolinio subjekto eisenoje nepastebėta jokių eisenos sustingimo požymių, kairės ir dešinės kojos motorinės operacijos atliekamos simetriškai, išskyrus kairės kojos kelio didžiausios amplitudės pagreičio kojos susilietimo ir pakilimo su žeme metu. Abiejų kojų susilietimo su žeme laiko ilgis simetriškas, kas ir patvirtina pusiausvyrą eiseną. Parkinsono subjektas parodė dešinės kojos eisenos sąstingio nesimetriškumą. Pacientas naudojo kompensacijos mechanizmą – jis peržengdavo dešinės pusės judesius, kuomet pradėdavo ėjimą. Kairės pusės judesiai atliekami tik palaikant balansą. Pati eiseną yra labai lėta, o sustingimas įvyksta tik dešinėje pusėje. Iš gautų rezultatų galima padaryti išvadą, kad analizuojamas pacientas visuomet bando kompensuoti eisenos pokytį, naudodamasis vieno kūno šonu. Tyrimas atliktas tik su vienu pacientu, todėl nėra galima tiksliai teigti, kad kompensacija vyksta tik iš dešinės pusės, tačiau ją galima tvirtai sakyti, kad naudojantis kompensacijos mechanizmu – egzistuoja žingsnio variacijos tikimybė. Tokią hipotezę sėkmingai patvirtina ir prieš tai atlikta darbo analizė [25]. Kompensacijos nebuvimas kontrolinio subjekto atveju garantuoja, kad žingsnio ilgio variacija nėra didelė, o atvirkščiai – artės į nulį.

**2.1 lentelė** Klasifikavimo taiklumas ir tikslumas, panaudojus dirbtinių neuronų tinklo klasifikatorių [1]

	Co	Pt
Tikslumas	0,950	0,950
Taiklumas	0,900	1,000

## 2.2. Dirbtinių neuronų tinklų metodas

Tolimesnis darbas apžvalgai yra “A ground reaction force artificial neural network classifier for the diagnosis of Parkinson’s disease” [1]. Duomenys analizei panaudoti tokie patys, kokie naudojami ir šiame darbe, iš tos pačios duomenų bazės, tačiau pagrindinis nagrinėjamo darbo skirtumas yra naudojamų subjektų skaičius – analizei pasirinkta 40 subjektų, turinčių Parkinsono ligą ir 40 subjektų, neturinčių neurologinių sutrikimų. Nėra argumentuota kodėl toks subjektų skaičius panaudotas analizei, kadangi duomenų bazėje yra 93 subjektai su Parkinsono ligos sutrikimais ir 73 kontroliniai, sveiki subjektai. Iš atliktos pirminės analizės pastebėta, kad Parkinsono subjektai vŽRJ signale turi mažiau galios aukštuose dažniuose, žemesnę pirmo ir antro maksimumo amplitudę ir pavėluotą pirmą maksimumą. Atlikta DFA (angl. *Detrended fluctuation analysis*) parodė, kad Parkinsono subjektai turi ilgesnį vidurinį ilgo laikotarpio žingsnio pakilimo nuo žemės fazės koreliacija.

Darbe panaudoti 8 signalo požymiai: vidutinė galia tarp 0,5 Hz ir 1,5 Hz (dB), vidutinė galia tarp 1,5 Hz ir 20 Hz (dB), kojos pakilimo nuo žemės santykis viso žingsnio atžvilgiu (procentais), pirmo maksimumo reikšmė (N/BW); antro maksimumo reikšmė (N/BW), kairės kojos atliktos DFA analizės skalės  $\alpha$  eksponentė, dešinės kojos atliktos DFA analizės skalės  $\alpha$  eksponentė [7]. Matmenų mažinimo metodas nepritaikytas. Klasifikavimui pasirinktas dirbtinių neuronų tinklas su 8 paslėptais neurais, apmokymas vyko naudojantis porinio mastelio gradiento atvirkštinį dauginimą (angl. *scaled conjugate gradient backpropagation algorithm*). Apmokymui pasirinkta po 20 subjektų iš kiekvienos grupės, dirbtinių neuronų tinklo tikrinimas vykdytas panaudojus po 10 subjektų iš kiekvienos grupės, testavimas atliekamas panaudojus po 10 subjektų duomenis iš kiekvienos grupės. Klasifikavimo taiklumas ir tikslumas yra pateikti 2.1 lentelėje. Klasifikatoriaus tikslumas yra 95 %, kontroliniai subjektai atpažinti su 90 % taiklumu (vienas iš dešimt subjektų atpažinti klaidingai), Parkinsono liga sergantys subjektai atpažinti su 100 % taiklumu, visi subjektai atpažinti teisingai.

Atliktas tyrimas pateikia gerus klasifikavimo rezultatus. Metodo trūkumas yra naudojamas dirbtinių neuronų klasifikatoriaus. Jis neleidžia įžvelgti kaip būtent yra apibendrinti naudojami požymiai požymių erdvėje. Taip pat nepavaizduota kaip esami duomenys pasiskirstę požymių erdvėje. Kadangi naudojami 8 požymiai (kas apibrėžia ir požymių erdvės matmenų skaičių), galima panaudoti matmenų mažinimo algoritmą ir pateikti esamus rezultatus grafiniu pavidalu, vizualinei analizei. Taip pat nurodyti požymiai yra skirtingos skalės. Kiekviena iš išvardintų

požymių gali turėti savo amplitudės diapazoną. Labai svarbu yra normuoti naudojamus signalo požymius taip užtikrinant kiekvieno požymio lygų įvertį klasifikatoriui apibendrinant duomenis. Šiuo metu nėra įmanoma nustatyti kiek vienas požymis turi įtakos bendram algoritmo atpažinimo veikimo efektyvumui.

### 2.3. Atraminių vektorių mašinos metodas

Tolimesnis darbas, pateikiamas apžvalgai “Statistical Analysis of Gait Rhythm in Patients With Parkinson’s Disease” [25]. Šis darbas nagrinėja galimas žingsnio variacijos galimybes ir pateikia signalo nuokrypio skaičiavimą kaip pagrindinį požymį, atskiriantį kontrolinius subjektus nuo sergančių. Nagrinėjime panaudoti žingsnio pakilimo nuo žemės ilgis, žingsnio prisilietimo prie žemės ilgis ir bendras žingsnio ilgis. Duomenys gaunami iš vertikalios žemės reakcijos jėgos matavimų. Autorių teigimu, žingsnio pakilimo ir nusileidimo ilgiai turi aukštą koreliacijos koeficientą su bendro žingsnio ilgiu, todėl nagrinėjime panaudotas tik kairės kojos žingsnio ilgis (2.2 lentelė). Gautas požymių vektorius apskaičiuotas, naudojantis “vieno palikimo-išmetimo tarpusavio patikros” (angl. *leave-one-out cross-validation*) metodu. Jis parodė, kad mažiausių šaknų atraminių vektorių mašina (angl. *Support Vector Machine*), su daugianariu branduoliu sugeba klasifikuoti duomenis su 90,32 % tikslumu.

Nagrinėjime naudoti duomenis, pateikti “Hausdorff et al” [6], kurie yra laisvai prieinami iš PhysioNet duomenų bazės [14]. Tokie duomenys yra panaudoti ir šiame darbe, todėl reikia įsigilinti ir į subjektus, kurie dalyvavo duomenų rinkimo metu. Tyrime dalyvavo 16 sveikų subjektų, 20 – 74 metų amžiaus. Kontroliniai subjektai yra visiškai sveiki neurologiškai, neturi jokių širdies ligų sutrikimų. Parkinsono subjektų skaičius yra 15, kurių amžius variavo nuo 44 iki 80 metų. Ūgis ir svoris žymiai nesiskyrė tarp kontrolinių ir Parkinsono subjektų, todėl į tokius duomenis dėmesys nebuvo kreipiamas. Parkinsono subjektai vertinami pagal “Hoehn and Yahr (HAY)” vertinimo sistemą, nustatant jų ligos stadiją. Du pacientai surinko 1,5 HAY balo (lengva stadija), keturi pacientai surinko tarp 2 ir 2,5 HAY balo, penki surinko 3 HAY balo (vidutinė stadija), likusieji surinko virš 3 HAY balų, jie turėjo žymesnių eisenos sutrikimų.

Remiantis gauta pasiskirstymo funkcija, kuri žymima  $\hat{p}(x_b)$ , kur  $x_b, b = 1, 2, \dots, B$ , kur  $B$  nusako segmentų skaičių, skaičiuojant  $x$  amplitudės ruožą, vidurkis  $\mu$  ir variacija  $\sigma^2$  apskai-

**2.2 lentelė** Koreliacijos koeficientai tarp  $\sigma$  bendro žingsnio ( $\sigma_r$ ), kojos pakilimo ( $\sigma_w$ ) ir kojos prisilietimo ( $\sigma_a$ ) prie žemės ilgio parametrų [25]

$\sigma$	$\sigma_r$	$\sigma_w$	$\sigma_a$
$\sigma_r$	1,00	0,99	0,94
$\sigma_w$	0,99	1,00	0,94
$\sigma_a$	0,94	0,94	1,00





**2.3 lentelė** Bendro žingsnio, kojos pakilimo, kojos prisilietimo prie žemės vidurkio, vidutinio nuokrypio duomenys [25]

Žingsnio fazė	Statistinis parametras	Kontrolinis subjektas	Parkinsono subjektas
		Vidurkis $\pm$ nukrypimas	Vidurkis $\pm$ nukrypimas
Bendras žingsnis	$\mu(s)$	$1,09 \pm 0,09$	$1,12 \pm 0,12$
	$\sigma(s)$	$0,03 \pm 0,01$	$0,06 \pm 0,04$
	$STC$	$12,44 \pm 10,46$	$55,87 \pm 29,89$
Kojos pakilimas	$\mu(s)$	$0,39 \pm 0,04$	$0,37 \pm 0,05$
	$\sigma(s)$	$0,02 \pm 0,01$	$0,04 \pm 0,02$
	$STC$	$8,5 \pm 9,39$	$54,13 \pm 26,96$
Kojos prisilietimas	$\mu(s)$	$0,70 \pm 0,06$	$0,75 \pm 0,09$
	$\sigma(s)$	$0,02 \pm 0,01$	$0,05 \pm 0,04$
	$STC$	$15,81 \pm 12,07$	$61,27 \pm 25,62$

čiuojama:

$$\mu = \sum_{b=1}^B x_b \hat{p}(x_b), \quad (2.1)$$

$$\sigma^2 = \sum_{b=1}^B (x_b - \mu)^2 \hat{p}(x_b). \quad (2.2)$$

Tyrimo metu nustatyta, kad Parkinsono subjektų žingsnio ilgio variacija yra žymiai padidėjusi, lyginant su kontroliniais subjektais (2.3 lentelė), tačiau žingsnio vidurkis lieka beveik nepakitęs. Lentelėje taip pat pateikiami signalo nuokrypio skaitiklio (angl. *Signal Turn Counter* –  $STC$ ) požymio vertės. Signalas  $x(i)$   $STC$  skaičiuojamas, remiantis tokiu loginiu ryšiu:

$$if \begin{cases} [x(i) - x(i-1)][x(i+1) - x(i)] < 0 \\ |x(i+1) - x(i)| \leq Th, 2 \leq i \leq I-1 \end{cases} \quad (2.3)$$

kur  $Th$  nusako ribinę vertę,  $I$  nusako signalo ilgį.

Toks požymis taip pat turi žymių skirtumų tarp kontrolinio subjekto ir Parkinsono subjekto, tačiau šis požymis turi neigiamą aspektą – jos rezultatas priklauso ne nuo signalo kitimo pobūdžio, o nuo pasirinktos ribinės vertės (toks matematinis procesas dar vadinamas “derinimu”), o tai neatspindi Parkinsono subjektų eisenos ypatybių. Tokio matematinio “derinimo” rezultate gali įvykti, kad algoritmas priderintas būtent prie nagrinėjime naudojamų signalų duomenų, tačiau jis gali būti visiškai bevertis į jį pateikus visiškai nematytus algoritmui duomenis.

Iš gauto tyrimo rezultato, galima spręsti, kad bendro žingsnio, kojos pakilimo, kojos prisilietimo prie žemės signalo ilgio dispersija ( $\sigma$ ) gali būti panaudotas kaip požymis, norint identi-

## 2.4 lentelė Pagrindiniai eisenos parametrai remiantis atlikta analize [12]

Žingsnio ypatybė	Kontrolinis subjektas	Parkinsono subjektas
	Vidurkis $\pm$ nukrypimas	Vidurkis $\pm$ nukrypimas
Žingsnio ilgis (m)	0,550 $\pm$ 0,080	0,420 $\pm$ 0,200
Eisenos greitis (m/s)	0,920 $\pm$ 0,160	0,650 $\pm$ 0,330

fikuoti Parkinsono subjekta.

## 2.4. Statistinės analizės metodas

Penktas ir paskutinis darbas apžvalgai yra gilesnės statistinės eisenos analizės darbas, “Statistical Analysis of Parkinson Disease Gait Classification using Artificial Neural Network” [12]. Darbe išnagrinėtos trys grupės eisenos požymių: pagrindinės, kinetinės ir kinematinės. Pagrindiniai požymiai: žingsnio laikas, žingsnių skaičius per minutę, žingsnio ilgis, eisenos greitis. Kinetiniai požymiai: maksimalus pėdos kontaktas su žeme pagal vertikale; maksimalus pėdos kontaktas su žeme pėdos pakilimo metu pagal vertikale; minimali jėga tarp dviejų kojos prisilietimo prie žemės maksimumų pagal vertikale; maksimali pėdos jėga pagal horizontale, kuomet pėda prisiliečia prie žemės; maksimali pėdos jėga pagal horizontale, kuomet koja pakyla nuo žemės. Kinematiniai požymiai: kulkšnies kampas, kuomet pėda prisiliečia prie žemės; kelio kampas, kuomet pėda prisiliečia prie žemės; klubo kampas kuomet koja prisiliečia prie žemės; maksimalus kulkšnies, kelio ir klubo kampas ėjimo metu. Duomenys tokiems požymiams apskaičiuoti surinkti pasitelkus slėgiui jautrų paviršių, kurio diskretizavimo dažnis yra 200 Hz ir 37 markerių, kurie buvo pritvirtinti prie subjektų kojų ir stebėti infraraudonųjų spindulių kamerų. Visi duomenys apdorojami *Vicon*<sup>®</sup> programinio paketo. Norint išvengti slėgiui jautriam paviršiumi surenkamų duomenų priklausomybės nuo subjektų svorio, duomenys normuojami pagal svorį:

$$Normuotas\ GRF(\%) = \frac{GRN(N)}{Svoris(N)} * 100\% \quad (2.4)$$

Iš viso duomenų surinkimo metu dalyvavo 12 Parkinsono liga sergančių subjektų ir 20 kontrolinių subjektų, kurie neturėjo savo ligos istorijoje jokių neurologinių susirgimų.

Klasifikavimo uždaviniui spręsti panaudotas dirbtinis neuronų tinklas su daugiasluoksniu perceptronu. Duomenis algoritmui pateikiami padalinus visus turimus duomenis į keturis rinkinius.

Darbo tikslas yra pirmiausiai analitiškai išanalizuoti visus gaunamus duomenis ir padaryti hipotezę, kokie eisenos požymiai geriausiai identifikuoja subjektų grupes. Vėliau, pagal klasifikatoriaus tikslumo ir taiklumo rezultatus, parinkti geriausius požymius, pagal kuriuos vyks subjektų atpažinimas. Pagrindiniai eisenos parametrai, kurie labiausiai skyrėsi tarp subjektų, yra pateikti 2.4 lentelėje. Pastebėta, kad kontrolinių subjektų žingsnio ilgis yra ilgesnis už

**2.5 lentelė** Dirbtinių neuronų tinklo klasifikavimo rezultatai, panaudojus skirtingas požymių grupes [12]

Požymių grupė	Klasifikavimo taiklumas
Pagrindiniai	81,25 %
Kinetiniai	81,25 %
Kinematiniai	84,38 %
Pagrindiniai ir kinetiniai	87,50 %
Pagrindiniai ir kinematiniai	87,50 %
Kinetiniai ir kinematiniai	84,38 %
Pagrindiniai, kinetiniai ir kinematiniai	87,25 %
Keturi didžiausio skirtumo požymiai	95,63 %

Parkinsono subjektų žingsnio ilgį. Iš pagrindinių požymių išskirtas ėjimo greitis. Parkinsono subjektų eisenos greitis pažymėtas kaip lėtesnis, lyginant su kontroliniais subjektais. Jėgos jutiklių parametrų palyginimas parodė, kad daugiausiai informacijos pateikia minimumas, kuris yra tarp pėdos pirmo prisilietimo prie žemės maksimumo ir pėdos pakilimo nuo žemės maksimumo. Iš kinematinų požymių, didžiausias skirtumas tarp subjektų yra klubo maksimalus kampas:  $19,904 \pm 8,623$  laipsnių kontrolinių subjektų atveju ir  $11,271 \pm 6,157$  laipsnių Parkinsono subjektų atveju.

Atlikus analitinį palyginimą, toliau atliktas požymių tikrinimas, remiantis klasifikatoriaus taiklumo rezultatais. Rezultatai yra pateikiami 2.5 lentelėje. Pagrindinių ir kinetinių požymių grupės surinko po vienodą klasifikavimo taiklumo rezultatą. Tai leidžia daryti išvadą, kad abi šių požymių grupės rodo panašias eisenos savybes. Kinematinų požymių taiklumo rezultatas yra 84,38 % ir yra didesnis už dviejų požymių grupių taiklumą. Sujungus pagrindines ir kinetines požymių grupes pasiektas didesnis klasifikavimo taiklumo rezultatas 87,50 %, tokį patį rezultatą pateikė ir pagrindinių su kinematiniais požymiais sujungimas. Kiti galimi požymių grupių sujungimai nepateikė geresnių klasifikavimo taiklumo rezultatų. Geriausią rezultatą pateikė keturių požymių grupė, kuri paaiškėjo po atliktos statistinės analizės: žingsnio ilgis, eisenos greitis, klubo maksimumo kampas ir minimumas tarp dviejų pėdos prisilietimo prie žemės maksimumų. Klasifikavimo taiklumas 95,63 %.

Iš aptarto darbo, šiame darbe yra panaudoti požymiai požymių erdvei sudaryti, bei pritaikytas dirbtinių neuronų tinklas, klasifikavimui atlikti.

## 2.5. Parkinsono ligos stadijos nustatymo metodas

Tolimesnė sistema apžvalgai siūlo jau konkretų algoritmą ir techninę įrangą jo įgyvendinimui – “Characterization of gait abnormalities in Parkinson’s disease using a wireless inertial sensor system” [21]. Darbe panaudoti žingsnio, siūbavimo ir nuokrypio sukimo fizinės eisenos

**2.6 lentelė** Klasifikavimo jautrumas, taiklumas, klaidingai klasifikuotas rodilis ir tikslumas, naudojant pastovų ir kintantį klaidingo klasifikavimo metodo kriterijų [21]

	Pastovus kriterijus	Kintantis kriterijus
Jautrumas	93,3 %	88,9 %
Taiklumas	95,8 %	100,0 %
Klaidingai klasifikuotas rodiklis	4,2 %	0,0 %
Tikslumas	97,7 %	100,0 %

**2.7 lentelė** Sunkios, lengvos ir kontrolinio subjekto eisenos sutrikimo klasifikavimo tikslumas ir taiklumas, identifikuojant lengvą ir sunkų Parkinsono ligos eigos atvejį [21]

	Klasės taiklumas	Klasės tikslumas
Sunkus PD eisenos sutrikimas	52,4 %	84,6 %
Lengvas PD eisenos sutrikimas	66,7 %	64,0 %
Kontrolinis subjektas	91,7 %	71,0 %

požymiai ėjimo metu, panaudota PCA požymių erdvės sudarymui ir SVM klasifikavimo mechanizmui įgyvendinti. Naudojama Micro-AHRS sensorinė platforma į kurią įeina 50 – g trijų ašių tiesinio pagreičio jutiklis ir 1200 °/s trijų ašių kampinio pagreičio jutiklis. Gautas modelis sėkmingai veikia 93 % tikslumu ir taiklumu. Naudojamas svoriui jautrus (angl. *cost-sensitive*) apmokymas tam, kad nustatyti kokie požymiai turi didesnę reikšmę formuojant požymių erdvę, bei atliekant klasifikavimą. Sprendimas atliktas, panaudojus pakankamą skaičių subjektų – 23 su Parkinsono ligos diagnoze ir 16 kontroliniai subjektai, kurie neturi jokių neurologinių susirgimų ligos istorijoje. Tyrimo metu taip pat nustatytas tikslas identifikuoti mažą ir didelį eisenos sutrikimo atvejus. Iš dalyvaujančių Parkinsono subjektų, 11 turi didelį eisenos sutrikimą, 12 turi mažesnę eisenos sutrikimą.

Tyrimo iš viso panaudoti 67 normuojami laiko srities požymiai, kurie yra svarbūs motorinei veiklai. Duomenys normuoti dėl skirtingų požymių amplitudžių srities. Normavimas yra labai svarbi procedūra, prieš pateikiant duomenis matmenų mažinimo metodui. Taip yra užtikrinama, kad naudojami požymiai turi vienodą įtakos faktorių, projektuojant naują matmenų erdvę [9]. Po matmenų mažinimo metodo pritaikymo, liko tik 11 požymiai, kurių bendra variacija yra 80,20 %. Parkinsono ligos eigos sunkumui nustatyti, prie bendros požymių erdvės pridėti dar papildomi keturi požymiai. Sprendimas, pateikti klasifikatoriui dar papildomus keturis matmenis, nėra tiksliai argumentuotas. Nėra nurodyta ar požymiai įterpti po matmenų mažinimo ar prieš tai. Jeigu tai atlikta po matmenų mažinimo, tai kelia abejonių naujai sukonstruota erdvė – sistema nėra gerai apibendrinta, jina yra “priderinta” prie esamų duomenų, todėl kelia abejonių ir pačio produkto lankstumas naujų, nematytų duomenų atžvilgiu.

Nepaisant neaiškumų dėl požymių erdvės – vektoriaus palaikymo mašinos klasifikatorius su



radialiniu Gauso branduoliu subendrina požymių erdvę. Klasifikatoriaus tikslumo ir taiklumo rezultatai yra pateikiami 2.6 lentelėje. Geriausias klasifikavimo tikslumas pasiektas, naudojant kintamą klaidos vertinimo kriterijų. Nurodytas klasifikavimo metodas atpažįsta Parkinsono ligą sėkmingai (100 % tikslumas).

Atliekant sudėtingesnę klasifikavimą tarp lengvos eisenos sutrikimo, sunkios eisenos sutrikimo Parkinsono ir kontrolinio subjekto, klasifikavimo tikslumas ir taiklumas gaunamas prastėnis. Sunkios eisenos sutrikimas atpažįstamas prastai, taiklumas rezultatas yra 52,4 %, lengvos eisenos sutrikimo klasifikavimo taiklumas siekia 66,7 %. Geriausiai atpažįstamas tik kontrolinės eilės subjektas, taiklumas yra 91,7 %.

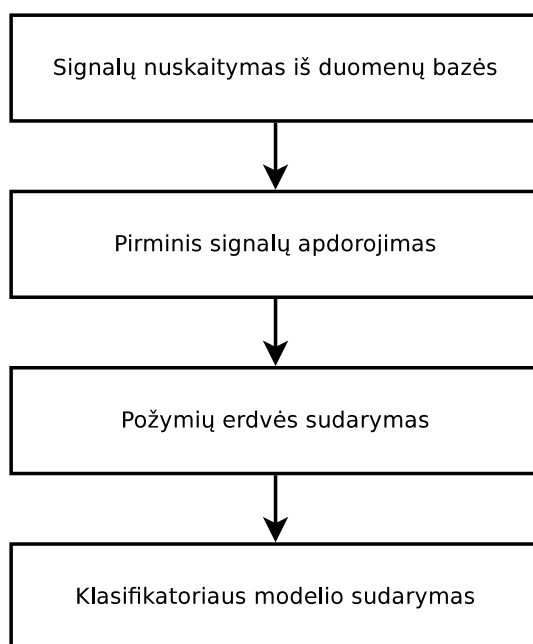
Aptartas metodas siūlo konkrečią įrangą, algoritmą ir požymius, tačiau užsibrėžtas tikslas pasiektas tik iš dalies – suprojektuota sistema, naudojanti pigius vertikalios pagreičio jutiklius, kuri atpažįsta Parkinsono ligą sergančių subjektų eigos sutrikimus, tačiau sistema labai blogai atpažįsta lengvą ir sunkų eisenos sutrikimą. Nurodyta kryptimi reikia atlikti didelį darbo kiekį, norint sėkmingai identifikuoti Parkinsono ligos stadiją ir atitinkamai pateikti HAY skalės vertinimą, analizuojant žmogaus eisenos požymius.

### 3. Signaļu analīzes programmas kūrimas

Šiame skyriuje apžvelgtos kylančios problemos kuriant signalų analizės programą, pateikti galimi jų sprendimo būdai. Skyrius prasidės bendros programos struktūrinės schemos sudarymo poskyriu. Jame pateikta bendra programos veikimo struktūra, nurodytas funkcinių modulių skaičius, kiekvieno nurodyto modulio paskirtis. Toliau aptarti galimi pirminio signalo apdorojimo etapai. Tai yra vienas iš svarbiausių programos žingsnių, kadangi nuo pirminio signalo apdorojimo kokybės priklauso ir visos programos rezultatas. Požymių išskyrimo poskyryje atliktas signalo požymių nagrinėjimas, jų analizė. Nagrinėjama požymių erdvė yra naudojama klasifikavimo metu, todėl būtina rasti tokius požymius, kurie skiriasi tarp kontrolinio ir Parkinsono subjekto. Požymio klasifikavimo programos kūrimo metu aptarti galimi naudoti klasifikatoriai, kiekvienam iš jų pateikti vienodi duomenys klasifikavimo užduočiai atlikti ir nuspręsta koks klasifikatorius geriausiai tinka nagrinėjamų požymių erdvei. Galiausiai, duomenų analizės programos kūrimo poskyryje pateiktas geriausiai užduotį atliekantis pirminis signalų apdorojimo metodas, matmenų mažinimo algoritmas, bei klasifikavimo mechanizmas.

### 3.1. Bendros programos struktūrinės schemos sudarymas

Priklausomai nuo taikomos metodikos, programos struktūrinės schemos skiriasi, kadangi vienas sprendimas reikalauja vieno tipo duomenų struktūros, kitas – kitos struktūros. Programos projektavimo metu išbandyti kelios programos versijos ir suprastinus naudojamas veikimo schemas, kiekviena iš programos versijų veikė pagal bendrą schemą, kurią galima apibrėžti 3.1 pav. Reikalaujama, kad kiekvienas iš algoritmų blokų būtų nepriklausomas nuo žemiau



### 3.1 pav. Pirminė programos schema



ar aukščiau einančio bloko, kas reiškia, kad “Pirminio signalų apdorojimo” bloko algoritmo pakeitimas neturi turėti jokios įtakos po jo einančiam “Požymių erdvės sudarymas” blokui.

Kiekvieno bloko užduotis:

1. “Signalų nuskaitymas iš duomenų bazės” bloko užduotis yra nuskaityti duomenis iš naudojamos “PhysioBank” duomenų bazės ir juos įkelti į Matlab aplinką. Kadangi duomenys pateikiami tekstiniu pavidalu, Matlab aplinkoje yra labai patogi funkcija tokiems duomenims nuskaityti – *dmlread*.
2. “Pirminis signalų apdorojimas” bloko užduotis yra pradinis signalų filtravimas, nuolatinės komponentės pašalinimas. Taip pat, į šį bloką įeina ir atskirų signalų išskyrimas iš bendrai gaunamo signalo: pėdos prisilietimas prie žemės signalas ir pėdos pėdos pakilimo nuo žemės signalas. Į bloką taip pat įeina ir slankiojančio lango algoritmo taikymas.
3. “Požymių erdvės sudarymas” bloko paskirtis yra pateikti pasirinktam klasifikatoriui išskirtus signalų požymius po pirminio signalo apdorojimo. Tokiais požymiais gali būti signalų dažninės komponentės (Furjė koeficientai), koreliacijos koeficientai tarp dešinės ir kairės kojos, savi-koreliacijos koeficientai dešinės ar kairės kojos. Požymiai detaliau nagrinėjami tolimesniuose poskyriuose. Į bloką taip pat įeina ir darbas su matmenimis – matmenų skaičiaus mažinimas ar naujos erdvės paieška, kurioje nagrinėjami duomenys yra geriau tiesiškai atskiriami.
4. “Klasifikatoriaus modelio sudarymas” bloko paskirtis yra klasifikatoriaus apmokymas ir jo testavimas. Klasifikatoriaus pasirinkimas yra labai svarbus klausimas darbe. Jis turi būti parinktas argumentuotai, išanalizavus kiekvieno naudojamo klasifikatoriaus tikslumo ir tikslumo rezultatus. Svarbus klasifikatoriaus aspektas yra požymių erdvės bendro dėsnio radimas arba erdvės aproksimacija. Šį faktą galima apibūdinti iš klasifikatoriaus rezultatų grafiko. Jeigu ilgą laiko tarpą į klasifikatorių yra siunčiami vienos klasės duomenys, o klasifikavimo rezultatai yra nestabilūs, ilgą laiką neturintis pastovaus rezultato, vadinasi, galima teigti, kad klasifikatorius blogai atliko erdvės duomenų pasiskirstymo dėsnio aproksimavimą, jis nėra tinkamas nagrinėjamai požymių erdvei.

Tolimesniuose poskyriuose aptarti įgyvendinti blokinės struktūros elementai.

### 3.2. Pirminio signalų apdorojimo programos kūrimas

Šiame skyriuje aptartas pirminio signalų apdorojimo programos kūrimas. Aptartas duomenų nuskaitymas iš duomenų bazės bylos tekstiniame pavidale, galimas signalo išskaidymas slankiojančio lango metodu arba signalo formos nuskaitymas, priklausomai nuo žinomo fizinio poveikio, kuriuo metu gautas signalas.

Programos kūrimo patogumui, parinkta direktorių architektūra:

- Pagrindinė direktorija



- <programos versija, nurodyta datos formatu>
- database
- cache

Kadangi programos kūrimo metu yra svarbu saugoti ankstesnes programos versijas, pasirinktas kasdienis programos versijos saugojimas: kiekvienos pradžioje darbas pradedamas su ankstesnės dienos kopija. Taip buvo išsaugotas kiekvienos darbo dienos programos versija ir taip galima peržiūrėti kokių analitinių kelių eita prie dabartinės programos versijos.

Duomenų nuskaitymas iš duomenų bazės vykdomas tekstiniu pavidalu. Iš viso duomenų byloje yra pateikiami 19 tipo duomenų, kurie aptarti užduoties analizės skyriuje. Matlab aplinkoje tekstiniu duomenų nuskaitymas yra atliekamas integruota *dlmread* funkcija. Funkcijos išvestyje yra pateikiama matrica, kuri susideda iš esamų duomenų byloje reikšmių. Kadangi duomenų apdorojimas vyksta analizuojant tik kairę koją, dėl to pritaikytas paprasta stulpelio išskyrimo operacija.

Slankiojančio lango algoritmas yra plačiai taikomas signalų apdorojimo sistemose. Algoritmo veikimo paaiškinimas toks – nagrinėjamas signalas  $\mathbf{X}(n)$ , kur  $n$  nusako laiko rodyklę. Nustatomas slankiojančio lango ilgis  $k$ , lango persidengimas  $p$ , momentinio laiko rodyklė  $i$ . Rodyklė  $i$  didinama tol, kol,  $i = k$ , tuomet turimas signalas yra saugomas laikinojoje atmintyje:

$$\mathbf{P}(0, k) = \{x(0), x(1), \dots, x(i)\}, \text{ kai } i \geq 3 \quad (3.1)$$

tuomet rodyklė  $i$  didinama toliau, kol rodyklės reikšmė netampa  $i = k + p$  ir gautas signalas saugomas laikinojoje atmintyje:

$$\mathbf{P}(1, k) = \{x(0 + p), x(1 + p), \dots, x(i + p)\}, \text{ kai } i \geq 3 \quad (3.2)$$

reikšmė  $i$  didinama toliau, kol negaunama reikšmė  $i = k + p * 2$  ir gautas signalas saugomas laikinojoje atmintyje:

$$\mathbf{P}(2, k) = \{x(0 + p * 2), x(1 + p * 2), \dots, x(i + p * 2)\}, \text{ kai } i \geq 3 \quad (3.3)$$

Apibendrintai slankiojančio lango metodą galima aprašyti viena formule:

$$\mathbf{P}(j, k) = \{x(0 + (p - 1) * j), x(1 + (p - 1) * j), \dots, x(i + (p - 1) * j)\}, \text{ kai } i \geq 3. \quad (3.4)$$

Toks sprendimas bus panaudotas ir įgyvendintas šiame darbe Matlab aplinkoje.

Kiekvienas gautas signalas visuomet turės nereikalingą triukšmą. Medicinoje naudojami įrenginiai turi atlikti savo užduotį kiek įmanoma geriau, todėl šiame darbe yra atliekamas





pradinis signalo dažnio filtravimas, norint pašalinti signalo nuolatinę komponentę ir aukšto dažnio triukšmą, kurio žmogaus kūnas negali generuoti. Filtrai aprašomi perdavimo funkcija:

$$H(s) = \frac{Y(s)}{X(s)}, \quad (3.5)$$

kur  $s = \delta + jw$ ,  $Y(s)$  išėjimo signalas,  $X(s)$  įėjimo signalas. Perdavimo funkcija nusako kiek pakis išėjimo signalas, prie tam tikro įėjimo signalo.

Matlab aplinkoje galima yra įgyvendinta daug tipo filtrų: *Butterworth*, *Chebyshev* 1 tipo, *Chebyshev* 2 tipo, *Elliptic*. Iš visų nagrinėjamų filtrų pasirinktas *Butterworth*, kadangi jo dažninė charakteristika yra lygiausia [23].

Sudėtingesnis signalo skaidymas už slankiojančio lango metodą yra skaidymas pagal žinomą fizinį aktyvumą. Dominantys signalai yra: kojos prisilietimo prie žemės signalas ir kojos pakilimas nuo žemės signalas. Algoritmo kūrimas pradedamas tikrinimo ar signalo vertė nėra lygi nuliui:

$$x(i) \in \mathbf{X}(i) \neq 0, \text{ kai } i \neq n \quad (3.6)$$

kur  $i$  yra laiko rodyklė,  $x(i)$  yra signalo  $\mathbf{X}(i)$  momentinė vertė,  $n$  yra signalo  $\mathbf{X}(i)$  ilgis. Jei-  
gu sąlyga tenkinama, momentinė signalo vertė yra perkeliama į laikinąją atmintį ir tikrinimas vyksta toliau. Tikrinimas sustabdomas, kuomet momentinė vertė yra lygi nuliui. Toliau tikrinamas signalo ilgis. Jei signalo ilgis yra didesnis už 200 momentinių verčių, skaitoma klaida. Ji gali įvykti duomenų sesijos rinkimo metu, kuomet subjektas nepakelia kojos per numatytą laiko tarpą. Jei signalo ilgis yra mažesnis už 10 verčių, skelbiama klaida. Toks greitas kojos pokytis nėra būdingas žmogaus eisenai. Klaidą gali sukelti aukšto dažnio triukšmas arba blogai prižiūrimas duomenų rinkimas sesijos metu. Signalui tenkinant visas sąlygas – nurodytas vektorius yra perkeliamas į laikinąją atmintį tolimesniam apdorojimui.

Turit mažiausiai du sąlygas tenkinančius kojos prisilietimo prie žemės signalus, galima tikrinti kojos pakilimo nuo žemės signalą. Remiantis žmogaus eisenos simetriškumo prielaida, kuri teigia, kad žmogaus kojos pakilimas ir prisilietimas prie žemės paviršiaus laikas turi būti tapatus – signalas tikrinamas pagal tokias pačias sąlygas, kaip ir kojos prisilietimo prie žemės signalas. Atlikus kontrolę, signalas perkeliamas į laikinąją atmintį tolimesniam apdorojimui. Aprašyta metodika įgyvendinta Matlab aplinkoje.

Laikinojoje atmintyje esant pėdos prisilietimo prie žemės signalui, tolimesnis signalo apdorojimas yra dviejų maksimumų išskyrimas ir vieno minimumo nustatymas. Pirmiausiai reikia įvesti pokyčio kitimo kriterijų  $\delta$ , pagal kurį bus sprendžiama kiek į neigiamą ašį signalas turi pakisti, apsprendžiant signalo maksimalią vertę. Analizuojant signalą  $X_p(n)$ , kur  $p$  yra pėdos prisilietimo prie žemės signalo skaičiaus rodyklė,  $n$  yra laiko rodyklė, o  $\mathbf{X}$  yra pats signalas,

pirmas maksimumas yra randamas tikrinant sąlygą:

$$if \begin{cases} x(i) - x(i-1) \geq \delta & 0 \\ x(i) - x(i-1) < \delta & x(i-1) \end{cases} \quad (3.7)$$

Tikrinimas vyksta, kol reikšmė yra nelygi nuliui, tuomet signalo reikšmė yra įrašoma į laikinąją atmintį. Antras maksimumas randamas ieškomas naudojant tą patį matematinį modelį, tačiau prieš tai apvertus signalą pagal laiko ašį. Signalo minimumas randamas tarp dviejų maksimumu lokalaus minimumo būdu:

$$if \begin{cases} x(i-1) - x(i) \geq 0 & 0 \\ x(i-1) - x(i) < 0 & x(i-1) \end{cases} \quad (3.8)$$

Tikrinimas nutraukiamas, kuomet reikšmė lygi ne nuliui. Aprašyta metodika įgyvendinta Matlab aplinkoje.

Pirminio signalų apdorojimo programa baigiasi nurodytais metodais. Sekantis žingsnis yra apdorotų duomenų perdavimas tolimesnei programos pakopai – požymių išskyrimui.

### 3.3. Požymių išskyrimo programos kūrimas

Požymių išskyrimas priklauso nuo pirminio signalo apdorojimo mechanizmo. Jeigu ankstesniame bloke pakeisime ribinį filtro dažnį ir kaip nagrinėjamą požymį yra parinkti koreliacijos koeficientai ar dažninės komponentės – požymio vektorius pakis, todėl prieš kiekvieną požymių analizavimo ciklą yra pateikta ir pirminiam signalų apdorojimo bloke naudojami metodai. Taip pat, kadangi požymių išskyrimo programos kūrimas taip pat sprendžia ir matmenų klausimą – skyriaus pradžioje trumpai apžvelgti populiariausi šiame darbe panaudoti matmenų mažinimo metodai.

#### 3.3.1. Matmenų mažinimas

Matmenų mažinimas suteikia dideles galimybes supaprastinti klasifikavimo uždavinį. Kuomet duomenys požymių erdvėje nėra tiesiškai atskirti, yra sunku analitiškai nustatyti ar klasifikatorius ras apibendrintą duomenų priskyrimo funkciją. Norint užtikrinti, kad klasifikatorius veiks teisingai – yra atliekamas matmenų praskyrimas. Metodas taip pat taikomas, kuomet norima sumažinti matmenų skaičių. Problema iškyla, kuomet norima įgyvendinti sukurta sprendimą įterptinėje sistemoje, kurioje negalima pasiekti didelių aparatinių resursų. Turint didelių matmenų duomenis taip pat kyla ir pačios sistemos kaina, dėl tų pačių aparatinių resursų reikalavimų. Norint išvengti tokių problemų – taikomi matmenų mažinimo algoritmai. Populiariausi matmenų mažinimo algoritmai yra:

- Tiesinė diskriminanto analizė (angl. *Linear Discriminant Analysis – LDA*);
- Principinė komponentių analizė (angl. *Principal Component Analysis PCA*).

Taip pat, kiekvienas iš paminėtų būdų taip pat gali būti papildytas branduolio (angl. *Kernel*) funkcija. Metodų apibūdinimas pradėtas nuo paprasčiausios iš dviejų – PCA.

Principinė komponentių analizė [18] yra statmena ir paprasta transformacija, kuri yra daug kur naudojama dėl paprastos matematinės pusės ir lengvo įgyvendinimo. Egzistuoja skirtingi PCA metodo įgyvendinimo algoritmai. Vienas iš jų yra vienetinės reikšmės skaidymas (angl. *Singular Value Decomposition* – SVD). Literatūroje PCA ir SVD dažniausiai minimi kaip sinonimai. SVD yra greitas, tačiau didelių atminties resursų reikalaujantis metodas. Kuomet dideli atminties resursai nėra prieinami (įterptinėje sistemoje), reikia naudoti kitą įgyvendinimo versiją – naudoti tikrinių vektorių skaidymą. Toks sprendimas užima didesnius laiko resursus, lyginant su SVD, tačiau jam reikia mažesnių atminties resursų ir jis tinka nagrinėti didelių matmenų duomenis. Toks sprendimas taikomas ir šiame darbe. PCA įgyvendinimas naudojant SVD ir tikrinių vektorių skaidymą yra pateiktas priede. Toliau seka PCA matematinis nagrinėjimas.

Matematiškai, PCA apibrėžiamas kaip statmena tiesinė transformacija, kuri transformuoja duomenis į naują koordinačių sistemą, kurioje didžiausią variaciją projektuojama į pirmą ašį, antroji didžiausia variacija (statmena pirmajai) projektuojama į antrą ašį. Procesas tęsiasi tol, kol yra pasiekiamas norimas matmenų skaičius.

Apibrėžta duomenų matrica  $\mathbf{X}^T$ , su empiriniu vidurkiu, kuris lygus nuliui (empirinis vidurkis reiškia vidurkio apskaičiavimą ir jo atėmimą iš duomenų), kur kiekviena eilutė  $n$  atspindi duomenų rinkinį, o stulpelis  $m$  – turimas duomenų matmenis. Matricos  $\mathbf{X}$  vienetinės reikšmės skaidymas išreiškiamas:

$$\mathbf{X} = \mathbf{W}\Sigma\mathbf{V}^T, \quad (3.9)$$

kur  $m * n$  matrica  $\mathbf{W}$  yra matricos  $\mathbf{X}\mathbf{X}^T$  tikrinių vektorių matrica,  $\Sigma$  matrica yra  $m * n$  stačiakampio įstrižainės matrica su realiais skaičiais įstrižainėje ir  $n * n$  tikrinių vektorių  $\mathbf{X}^T\mathbf{X}$  matrica  $\mathbf{V}$ . PCA transformacija, kuri sukonstruoja naujas matmenis, yra apibrėžiama:

$$\mathbf{Y}^T = \mathbf{X}^T\mathbf{W} = \mathbf{V}\Sigma^T\mathbf{W}^T\mathbf{W} = \mathbf{V}\Sigma^T. \quad (3.10)$$

Kadangi  $\mathbf{W}$  yra statmena matrica, kiekviena  $\mathbf{Y}^T$  eilutė yra matricos  $\mathbf{X}^T$  eilutės sukimas. Pirmasis  $\mathbf{Y}^T$  stulpelis yra pirmosios komponentės rezultatas, antras stulpelis yra antrosios komponentės rezultatas. Kiek stulpelių  $\mathbf{Y}^T$  turi, tiek ir rezultatų (matmenų) yra po transformacijos. Matlab aplinkoje

Tiesinė diskriminanto analizė [22] yra matmenų mažinimo metodas (kuris kartu yra naudojamas kaip ir klasifikatorius), yra vienas iš metodų, kuris neša kartu ir duomenų žymėjimo informaciją. Tai reiškia, kad mažinant matmenų skaičių, metodas turi žinoti kokie duomenys

priklauso kokiai duomenų klasei. Pagrindinis LDA tikslas yra didinti sekantį kriterijų:

$$\mathfrak{J}(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}, \quad (3.11)$$

kur  $S_B$  yra “išorinės klasės scatter matrica”,  $S_W$  yra “vidinė klasės scatter matrica”. Scatter matricų apibrėžimas yra:

$$S_B = \sum_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \quad (3.12)$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (3.13)$$

Branduolio metodas yra labai lengvai paaiškinamas pavyzdžiu [8]. Tarkim, egzistuoja tam tikri empiriniai duomenys:

$$(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}. \quad (3.14)$$

Čia,  $\mathcal{X}$  yra netuščia sritis, iš kurios imami  $x_i$  duomenys. Tikslu arba žymekliu vadinami yra  $y_i \in \mathcal{Y}$  (sistemos atsakas),  $i, j \in [n]$  žymimas eilės, identifikavimo numeris, kur  $n := \{1, \dots, n\}$ .

Reikia pastebėti, kad sričiai  $\mathcal{X}$  nėra suteikta jokių apribojimų. Norint išspręsti apmokymo problemą reikia papildomos struktūros. Mašiniame apmokyme siekiama apibendrinti neturimus duomenis. Binarinio klasifikavimo atveju, turint naujus duomenis  $x \in \mathcal{X}$ , norima nuspėti jų žymeklį  $y \in \{\pm 1\}$ . Paprastai kalbant, norima pasirinkti tokį  $y$ , kuris labiausiai atspindėtų apmokymo metu naudotas  $(x, y)$  poras. Tam įgyvendinti reikalingi kuo panašesnis duomenys  $\mathcal{X}$  ir  $\mathcal{Y}$ . Pastarąjį palengvina tai, kad spėjimas gali būti arba identiškasis arba atvirkščias. Formaliai yra reikalinga tokia funkcija:

$$k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}, \quad (x, x') \rightarrow k(x, x'), \quad (3.15)$$

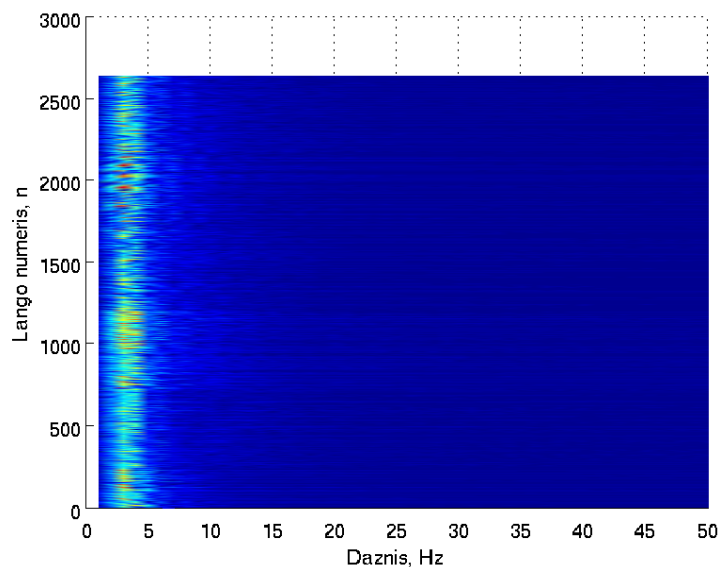
kuri tenkinta tokias sąlygą visiems  $x, x' \in \mathcal{X}$ :

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle, \quad (3.16)$$

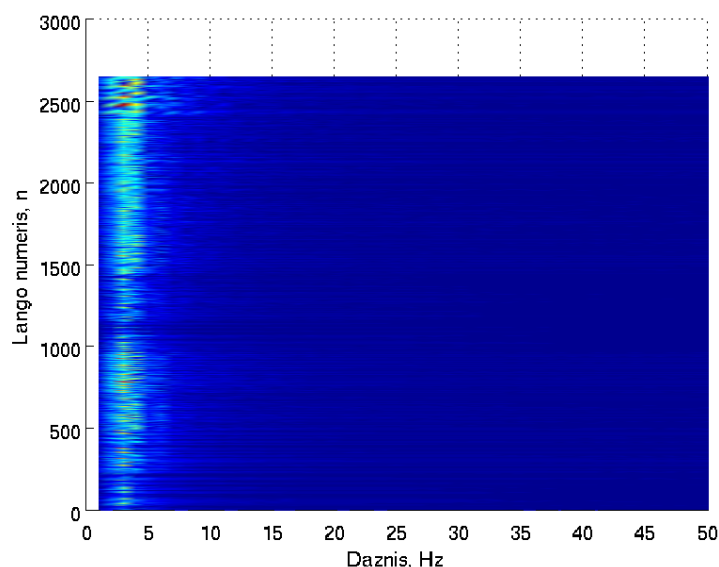
kur  $\Phi$  projektuota taškus į Hilberto plokštumą  $\mathcal{H}$ , vadinama požymių erdve. Vienodumo matmuo  $k$  dažniausiai vadinamas branduoliu, o  $\Phi$  vadinama požymių projekcija.

### 3.3.2. Požymių tyrimas

Šiame skyriuje išnagrinėti galimi signalų požymiai, pagal kuriuos galima atskirti Parkinsono liga sergantį subjektą nuo sveiko subjekto. Kaip minėta anksčiau, ne visus požymius galima apskaičiuoti, naudojantis vienu pirminio signalo apdorojimo mechanizmu, todėl kiekvieno požymio nagrinėjimo pradžioje yra paminėta ir pirminio signalo apdorojimo bloko sudėtis.



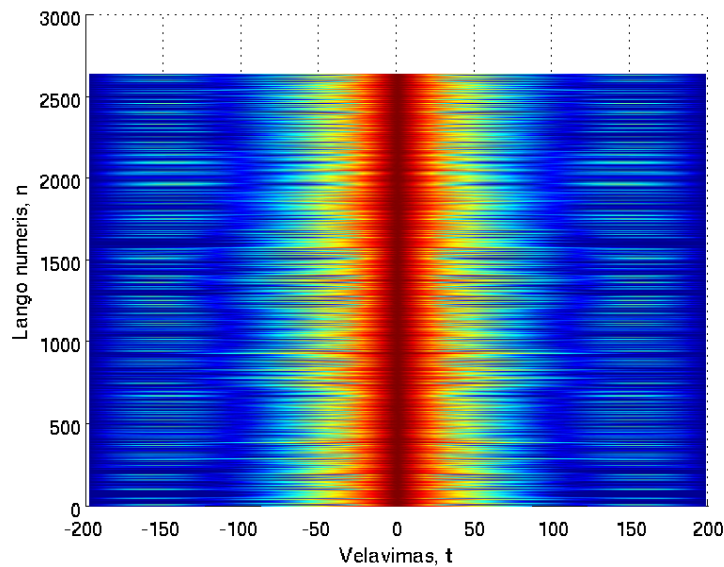
**3.2 pav.** Kontrolinių subjektų kairės kojos žingsnių signalų dažninės komponentės



**3.3 pav.** Parkinsono liga sergančių subjektų kairės kojos žingsnių signalų dažninės komponentės

Galimų požymių analizavimas gali būti pradėtas nuo signalo dažninių komponentių (Furjė transformacijos). Kaip nurodo šaltiniai [24], Parkinsono liga sergančių subjektų žingsniai turi statesnius šlaitus, kas iš signalų apdorojimo srities reiškia, kad signalas turi turėti aukštas dažnines komponentes. Šiam požymiui išskirti pirminiam signalo apdorojimo bloke panaudotas slankiojančio lango metodas. Taip kiekvienas signalas turės fiksuotą ilgį ir taip galima lyginti signalus tarpusavyje.

Toliau seka išanalizuotos Parkinsono liga sergančių subjektų ir kontrolinių subjektų žingsnių signalų dažninės komponentės. Kontrolinių subjektų kairės kojos žingsnių dažninės komponentės



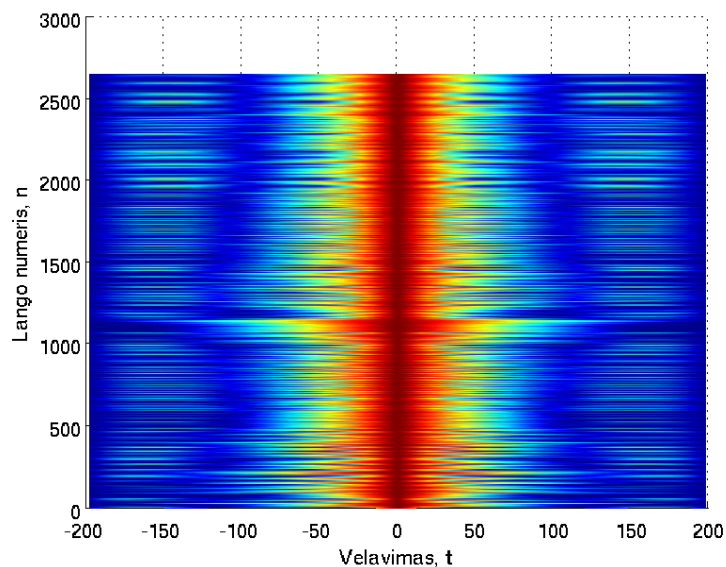
**3.4 pav.** Kontrolinių subjektų kairės kojos žingsnių signalų savi-koreliacijos koeficientai

tės yra pateiktos 3.2 pav. Parkinsono liga sergančių subjektų kairės kojos žingsnių dažninės komponentės yra pateiktos 3.3 pav. Kaip matoma iš duotų komponentių grafikų – tiek kontrolinių subjektų, tiek Parkinsono liga sergančių subjektų pagrindinės dažninės komponentės išsidėsto iki 5 Hz ruože. Ties 0 Hz dažnių komponentių nėra, kadangi jos pašalintos filtro pagalba. Už 10 Hz ribos, komponentės neneša visiškai jokios informacijos. Iš to galima padaryti išvadą, kad tiek Parkinsono liga sergančių subjektų, tiek kontrolinių subjektų eisenos yra visiškai vienodos dažnių srityje ir vien remiantis šita informacija nėra galima nustatyti ar subjektas serga Parkinsono liga ar ne.

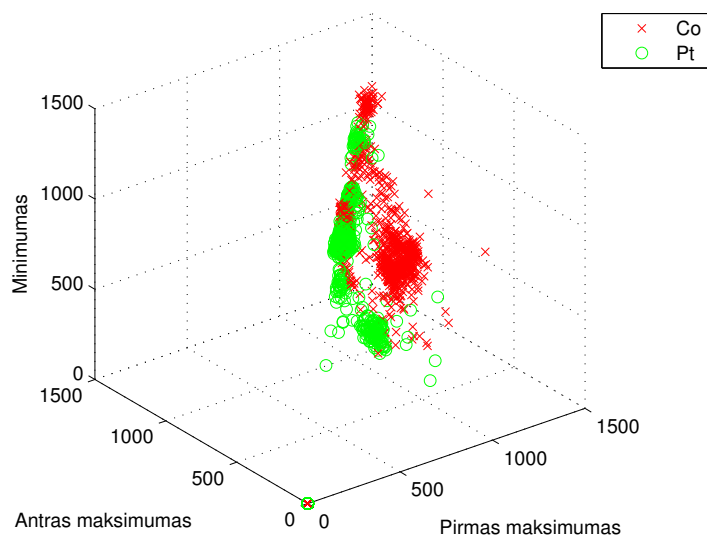
Tolimesnė analizė gali būti atlikta, remiantis koreliacijos koeficientais. Analizuojant šiuos požymius, galima remtis tokiu pačiu pirminiu signalo apdorojimo bloku, kaip ir analizuojant dažnines komponentes. Dėl to, kad esama tikrais kairės kojos signalai, galima taikyti tik savi-koreliacijos koeficientus. Narinėjamas požymis parodė gerus klasifikavimo rezultatus ankstesniame tyrime [15], kuriame, remiantis tiesinio pagreičio ir kampinio pagreičio jutiklių parodymais, reikėjo suprojektuoti algoritmą, gebantį atskirti tokias žmogaus veiklas: stovėjimas,ėjimas,ėjimas aukštyn laiptais,ėjimas žemyn laiptais.

Kontrolinių subjektų kairės kojos savi-koreliacijos koeficientai parodyti 3.4 pav. Parkinsono liga sergančių subjektų kairės kojos savi-koreliacijos koeficientai parodyti 3.5 pav. Kaip matosi iš korelogramos, tiek sergantys, tiek kontroliniai subjektai turi panašias, o kai kuriais atvejais ir tokias pačias, koreliacijos reikšmes. Remiantis vien tik turima informacija, nustatyti ar subjektas serga Parkinsono liga ar ne, nėra įmanoma.

Sudėtingesnė analizė seka iš dviejų globalių maksimumų ir vieno lokalaus minimumo požymių erdvės. Šie trys požymiai sudaro trijų matmenų plokštumą, kurią galima lengvai pavaizduoti. Pilna požymių erdvė pavaizduota 3.6 pav. Kaip matosi iš duotos erdvės, duomenys



**3.5 pav.** Parkinsono liga sergančių subjektų kairės kojos žingsnių signalų savi-koreliacijos koeficientai

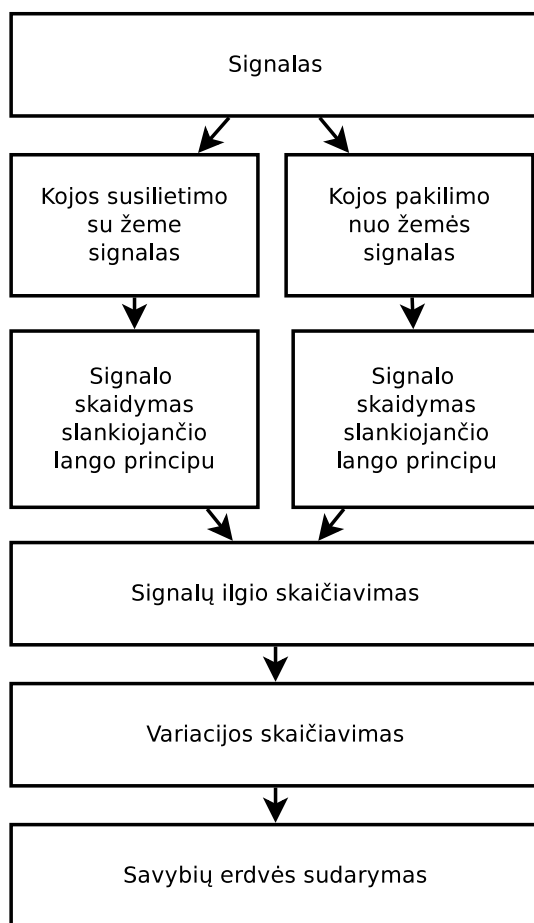


**3.6 pav.** Dviejų globalių maksimumų ir vieno lokalaus minimumo požymių erdvė

neturi jokio koncentracijos centro. Erdvėje jie pasiskirstę pagal nežinomą dėsnį.

Tokių duomenų pateikti klasifikavimui nėra galima. Grafike “Co” taškai, pažymėti kryžiumi, parodo kontrolinį subjektą, “Pt” taškai, pažymėti apskritimu, parodo Parkinsono liga sergančius subjektus.

Sekantys požymiai nagrinėjimui yra siūlomos atliktų tyrimų [5, 11, 4]. Nagrinėjamas požymis yra žingsnio fazės laiko variacija. Kojos susilietimo su žeme laiko variacija ir kojos pakilimo nuo žemės laiko variacija.

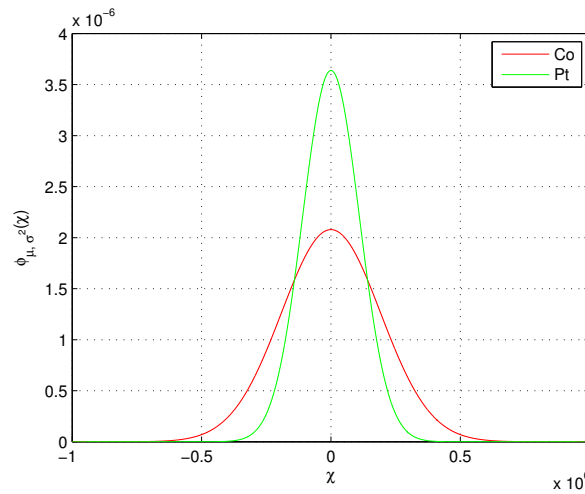


**3.7 pav.** Pirminis signalo apdorojimas, išskiriant kojos susilietimo su žeme laiko ir kojos pakilimo nuo žemės laiko variacijos požymiai

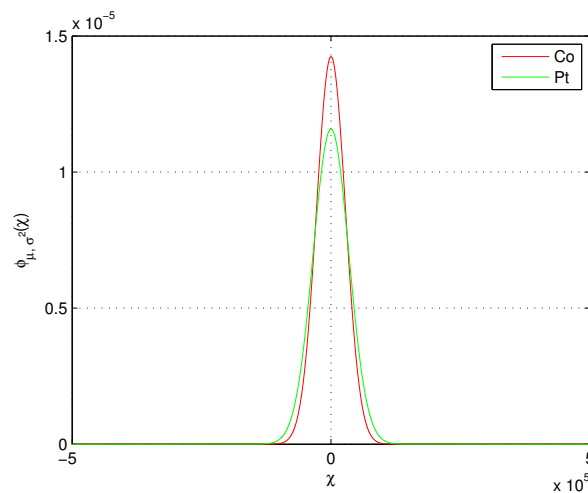
Išskiriant variacijos požymį iš signalo, reikalinga pakeisti pirminio signalo apdorojimo algoritmą. Kadangi kojos pakilimas nuo žemės ir kojos susilietimas su žeme gali būti vienas nuo kito nepriklausomi (subjektas sustojo ir stovi dviem kojomis ant žemės arba subjektas stovi tik ant kairės kojos), pirminis signalo apdorojimas turi būti atliekamas lygiagrečiai, t.y. tuo pat metu išskiriamas tiek kojos pakilimo nuo žemės signalas, tiek ir kojos susilietimas su žeme signalas. Tokios sistemos struktūrinis grafikas yra parodytas 3.7 pav. Sekantis žingsnis yra signalo išskaidymas slankiojančio lango principu. Abiejų signalų slankiojančio lango ilgis parinktas 4 verčių ilgio, su 2 verčių perdanga. Kuomet tiek kojos susilietimo su žeme, tiek kojos pakilimo nuo žemės keturi signalai patenkina keliamus reikalavimus – iš jų yra apskaičiuojami signalų ilgiai. Vėliau, abiejų etapų signalams yra paskaičiuojama jų variacija ir sudaroma požymių erdvė. Toliau nagrinėjamos abiejų signalų variacijos vieno matmens plokštumoje.

Abiejų signalų variacijų pasiskirstymai yra parodyti 3.8 pav. ir 3.9 pav. Abu pasiskirstymai turi vieną trūkumą – jų vidurkiai sutampa, tačiau gera žinia yra tai, kad jų variacijos skirtingos. Norint išspręsti iškilusią problemą, reikia taikyti matmenų praskyrimo metodus. Darbo metu išbandyti tokie matmenų praskyrimo metodai:





**3.8 pav.** Kojos prisilietimo prie žemės laiko variacija

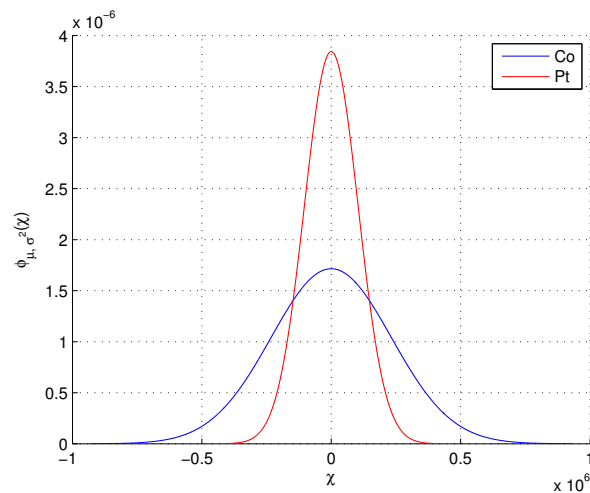


**3.9 pav.** Kojos pakilimo nuo žemės laiko variacija

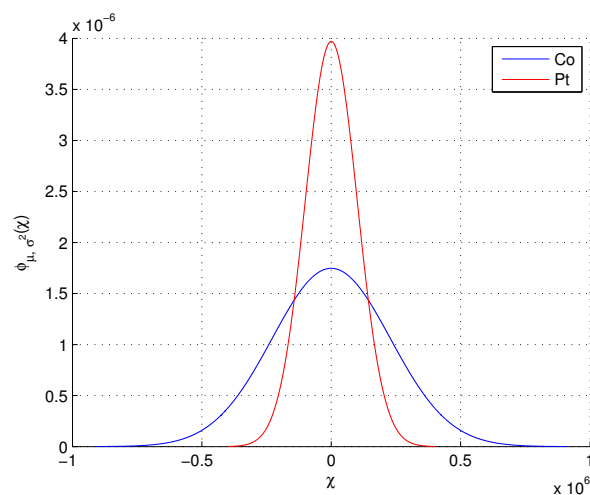
- PCA;
- LDA;
- Kernel PCA (su daugianariu, Gauso branduoliais);
- Kernel LDA (su daugianariu, Gauso branduoliais);

Vienmatė tiesinė PCA transformacija pavaizduota 3.10 pav. Vizualiai įvertinus gaunamą grafiką – vidurkis nepasikeitė, tačiau pakito variacija. Vidurkio pokyčio nėra pastebima, todėl teikti duomenis klasifikavimui nėra prasmės, kadangi to pačio vidurkio duomenis atskirti nėra įmanoma.

Sekantis metodas yra LDA. Transformacijos rezultatas yra pavaizduotas 3.11 pav. Vizualiai įvertinus gaunamą grafiką – tiek po LDA, tiek po PCA duomenų pasiskirstymas nėra gerai atskirtas. Galima daryti hipotezę, kad tiesinis duomenų atskyrimas šiuo atveju naudoti nėra



**3.10 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po PCA transformacijos



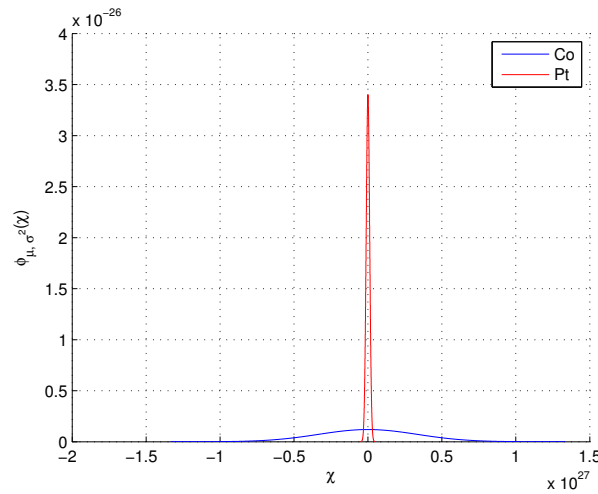
**3.11 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po LDA transformacijos

tinkamas. Reikia papildyti esamus algoritmus branduolio metodu ir papildyti transformacijas daugianariu arba Gauso branduoliu.

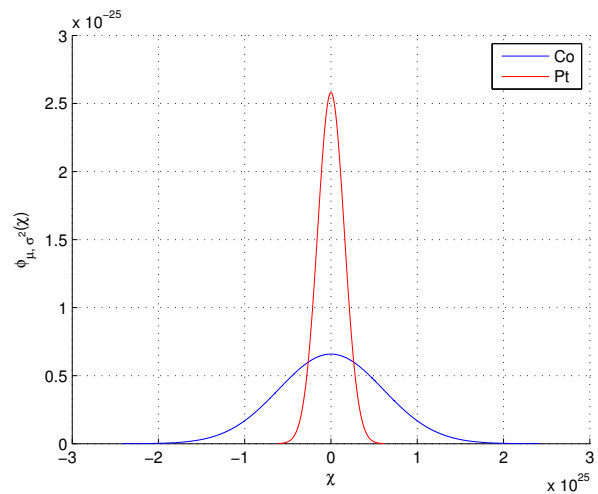
Branduolio metodas pritaikytas PCA transformacijai yra pavaizduotas 3.12 pav. ir 3.14 pav. Branduolio metodas pritaikytas LDA transformacijai yra pavaizduotas 3.13 pav. ir 3.14 pav. Daugianaris branduolys abiem atvejais panaudotas su 4 branduolio argumentu.

Vizualiai įvertinus gaunamą grafiką, panaudojus polinimonį branduolį, tiek PCA (3.12 pav.), tiek LDA (3.13 pav.) atveju – vidurkis iš savo vietos nepajudėjo. Abidvi transformacijos pakeitė variaciją. Tai nėra norimas tikslas, todėl toks branduolys nėra tinkamas. Panaudojus Gauso branduolį rezultatas labai pagerėja LDA transformacijos atveju (3.15 pav.), PCA su tokiu branduoliu (3.14 pav.) duomenis tik dar labiau suvienodina.

Remiantis pateikta analize galima teigti, kad geriausiai duomenis požymių erdvėje atskiria



**3.12 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po PCA transformacijos, naudojant daugianarį branduolį



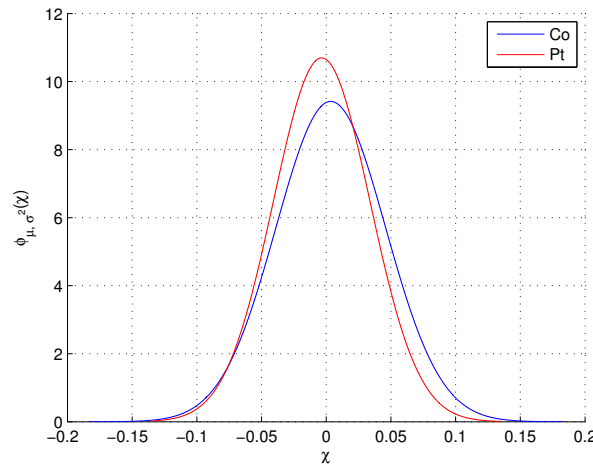
**3.13 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po LDA transformacijos, naudojant daugianarį branduolį

LDA su Gauso branduoliu. Tolimesniame darbe duomenis į klasifikatorių pateikiami po tokio tipo transformacijos.

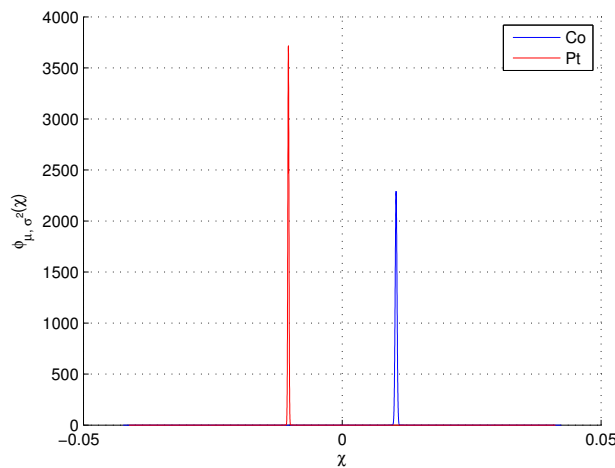
### 3.4. Požymių klasifikavimo programos kūrimas

Šiame skyriuje išnagrinėti ir pritaikyti populiariausi šiuo metu naudojami klasifikatorių metodai. Tokie klasifikatoriai yra [9]:

- Atraminių vektorių mašina (angl. *Support Vector Machine – SVM*);
- Paslėptas Markovo modelis (angl. *Hidden Markov Model – HMM*);
- Naivusis Bayes klasifikatorius (angl. *NayveBayes*);
- Tiesioginio sklidimo neuronų tinklas (angl. *Feed-Forward Neural Network – FFNN*);



**3.14 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po PCA transformacijos, naudojant Gauso branduolį



**3.15 pav.** Kojos prisilietimo prie žemės ir pakilimo nuo žemės laiko variacijos pasiskirstymas po LDA transformacijos, naudojant Gauso branduolį

Labai svarbu yra atskirti duomenis, kuriais klasifikatorius yra apmokamas ir kuriais jis yra testuojamas. Jeigu klasifikatoriaus apmokymo duomenys yra pakankamai apibendrinti, tuomet naujus duomenis klasifikatorius turėtų gerai atpažinti. Pateikiant klasifikatoriaus testavime tokius pačius duomenis, kaip ir apmokyme – atliekamas tikrinimas ar klasifikatorius teisingai “suprato” nagrinėjamus duomenis, tačiau tai neapibrėžia kiek gerai jis apdoros naujus duomenis.

Nurodytų klasifikatorių veikimas vertinamas taiklumo ir tikslumo parametrais. Taiklumas apskaičiuojamas:

$$\text{Taiklumas} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}, \quad (3.17)$$

### 3.1 lentelė Vektoriaus palaikymo mašinos tikslumas ir taiklumas

	Co	Pt
Tikslumas	0,489	0,489
Taiklumas	0,234	0,744

kur  $T_P$  – teisingai identifikuotų klasių skaičius,  $T_N$  – teisingai atmestų klasių skaičius,  $F_P$  – klaidingai identifikuotų klasių skaičius ir  $F_N$  – klaidingai atmestų klasių skaičius.

Tikslumas apskaičiuojamas:

$$Tikslumas = \frac{T_P}{T_P + F_P} \quad (3.18)$$

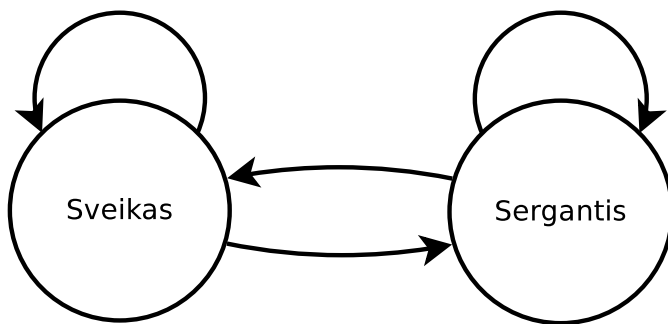
Visos tolesnės klasifikatorių patikros atliekamos remiantis tokia schema: iš turimų duomenų išskirta kojos prisilietimo ir pakilimo nuo žemės laiko ilgio variacijos požymiai. Visi duomenys padalinti į lygias tris dalis: matmenų operacijai atlikti duota 500 reikšmių, klasifikatoriaus apmokymui tolimesnės eilės 500 reikšmių, klasifikatoriaus testavimui tolimesnės eilės 500 reikšmių. Klasifikavimas atliekamas realiu laiku, t.y. klasifikatoriui pateikiami duomenis apie žingsnio stadijos variaciją po transformacijos ir klasifikatorius pateikia savo spėjimą.

Vektoriaus palaikymo mašina [2] šiuo metu yra populiariausias klasifikatorius nagrinėjant netiesiškai atskiriamus duomenis. Įgyvendinimas Matlab aplinkoje panaudotas iš [3]. Klasifikatoriaus testavimo metu naudojami tokie parametrai:

- SVM tipas – Sequential Minimal Optimization;
- Branduolio tipas – tiesinis;
- Sureguliuavimo konstanta – 2;
- Branduolio argumentas – 2;

Lentelėje 3.1 pateikti SVM tikslumo ir taiklumo duomenis. Kaip matosi, klasifikatorius veikia nepakankamai gerai – tikslumo koeficientas nėra didesnis negu pusę, abiem atvejais tik 0,489. Pirmos klasės atpažinimo taiklumas tėra 0,234, antros klasės atpažinimo taiklumas yra 0,744, tačiau to nepakanka. Iš padarytos klasifikatoriaus veikimo analizės galima teigti, kad klasifikatorius veikia blogai ir jo naudoti sprendime nėra galima.

Paslėptas Markovo modelis [16] yra vienintelis iš šiame darbe nagrinėjamų klasifikatorių, kuris turi laikinę informaciją. Tokia savybė suteikia “inkaro” galimybę – klasifikatorius gali užsilaikyti prie vienos klasės net ir tuomet, kai pagal požymių erdvę turi būti kita duomenų grupė. Tokia klasifikatoriaus savybė pritaikyta ankstesniame darbe, sudarant žmogaus eisenos atpažinimo sprendimą [15]. Pasirinktas modelis, susidedantis iš dviejų būsenų – “Sveikas”, “Sergantis”. Būsenos tarpusavyje yra sujungtos (3.16 pav.). Perėjimo tikimybės tarp modelio elementų parinktos žymiai mažesnės už tikimybę likti toje pačioje būsenoje. Įgyvendinimas Matlab aplinkoje panaudotas iš [10].



### 3.16 pav. Ergodinis paslēptas Markovo modelis

### 3.2 lentelė Paslėpto Markovo modelio tikslumas ir taiklumas

	Co	Pt
Tikslumas	0,500	0,500
Taiklumas	0,000	1,000

### 3.3 lentelė Naivaus Bayes klasifikatoriaus tikslumas ir taiklumas

	Co	Pt
Tikslumas	0,508	0,508
Taiklumas	0,764	0,252

Lentelėje 3.2 pateikti tikslumo ir taikumo duomenys. Kaip matome iš tikslumo rezultato – klasifikatorius teisingai priskiria tik pusei reikiamų duomenų. Iš to galime teigti, kad klasifikatorius yra labai blogai apmokytas ir visiškai nesugeba apibendrinti turimų duomenų. Klasifikatorius visus pateikiamus testavimo duomenis priskiria vienai klasei.

Naivus Bayes [17] yra vienas iš pirmųjų statistinių metodų paremtu klasifikavimo mechanizmas. Jis veikia labai paprastai – ieškoma tiesinės funkcijos, kuri geriausiai atskiria nagrinėjamus duomenis ir vieni žymenis priskiriami, jeigu duomenys yra vienoje linijos pusėje, atvirkšti žymenis priskiriami, jeigu duomenys yra kitoje linijos pusėje.

Lentelėje 3.3 pateikti tikslumo ir taiklumo duomenys. Tikslumo koeficientas viršija pusę, 0,508, tačiau tai yra mažai. Taiklumo koeficientas pirmuoju atveju yra neblogas, 0,764, tačiau antruoju atveju koeficientas yra visiškai nepatenkinamas, viso 0,252. Iš turimų rezultatų galima teigti, kad klasifikatorius veikia blogai ir duomenis vienmatėje erdvėje jis klasifikuoti teisingai negali.

Paskutinis klasifikatorius, kuris pritaikytas turimiems duomenims – tiesioginio sklidimo neuronų tinklas. Tai yra dirbtinių neuronų tinklų klasifikatorius, kuris veikia panašiai kaip ir Naivus Bayes – jis ieško funkcijos (tiesinės arba daugianarės), kuri geriausiai atskiria turimus duomenis. Bandymo metu pasirinktas vienas įėjimas, du išėjimai ir vienas paslėptas sluoksnis.

Lentelėje 3.4 pateikiami tiesioginio sklidimo dirbtinių neuronų tinklų tikslumas ir taiklumas. Kaip matosi iš rezultatų – klasifikatoriaus tikslumas pirmos klasės atžvilgiu yra 0,282,



**3.4 lentelė** Tiesioginio sklaidimo dirbtinių neuronų tinklų tikslumas ir taiklumas

	Co	Pt
Tikslumas	0,282	0,287
Taiklumas	0,000	0,392

taiklumas 0,000, antros klasės atžvilgiu klasifikavimo tikslumas yra artimas pirmai 0,287, tačiau turi geresnį taiklumą 0,392. Iš turimos patikros rezultatų galima spręsti, kad tiesioginio sklaidimo neuronų tinklas užduotį atlieka blogai.

Iš pateiktos analizės galima spręsti, kad realiu laiku pateikti klasifikavimo tikslumas ir taiklumas nėra geri, norint atlikti kokybišką spėjimą ligos atžvilgiu. Vektoriaus palaikymo mašinos tikslumas neviršija pusės, paslėptas Markovo modelis nesugeba atlikti tikslingo apmokymo, naivus Bayes klasifikatorius užduotį atlieka geriausiai iš visų nagrinėjamų klasifikatorių, tiesioginio sklaidimo neuronų tinklas užduotį atlieka blogiausiai, iš visų klasifikatorių, kurie sugebėjo bent kažkiek apibendrinti duodamus duomenis apmokymo metu. Atlikta analizė reikalauja kito būdo klasifikavimui atlikti. Naudojamas būdas aprašytas sekančiame poskyryje.

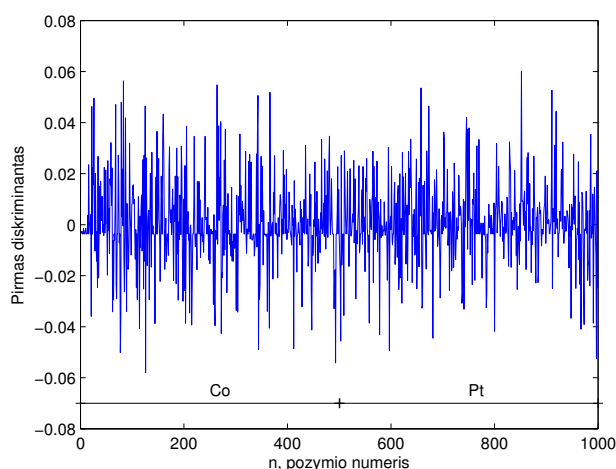
### 3.5. Duomenų analizės programos kūrimas

Šiame poskyryje aptartas duomenų analizės programos kūrimas. Pasinaudojus prieš tai esančių skyrių informacija yra pateiktas sprendimas, kuris leidžia efektyviai kiek įmanoma efektyviau atpažinti subjektų tipus.

Poskyryje “Požymių išskyrimo programos kūrimas” atlikta požymių ir matmenų mažinimo metodų analizė. Geriausias požymis, kuri atskiria kontrolinį subjektą nuo Parkinsono subjekto yra kojos pakilimo ir prisilietimo prie žemės laiko variacija. Geriausiai matmenų mažinimo klausimą išsprendė tiesinė diskriminanto analizė, naudojant Gauso branduolį, tačiau iškilo klasifikavimo problema – klasifikuojant duomenis realiu laiku, klasifikavimo rezultatas nepatenkinamas.

Pagrindinė problema, kodėl joks nagrinėjamas klasifikatorius neatliks korektiško klasifikavimo, tai dėl jam pateikiamų duomenų. Kaip pavyzdys yra pateikiamas klasifikatoriaus testavimo metu naudoti duomenis, pavaizduoti 3.17 pav. Duomenis yra pateikiami po atliktos transformacijos, todėl jie yra vieno matmens. Horizontalėje yra pateiktas požymio eilės numeris, vertikalėje – pirmas LDA diskriminant. Pirmi 500 požymiai priklauso kontroliniam subjektui, nuo 501 iki 1000 požymiai priklauso Parkinsono subjektui. Lyginant matomus duomenis su jų pasiskirstymu 3.15 pav., jie atrodo chaotiški. Iš vaizdinės analizės nėra įmanoma spėti kuri signalo dalis kuriai subjektų grupei priklauso.

Tokios išvados verčia projektuoti kitą klasifikavimo mechanizmą, kuris duomenis klasifikuotų ne kiekvieną pagal kiekvieną nagrinėjamą požymį, o pagal požymių grupę – į laikiną atmintį yra rašoma požymio reikšmė, laukiama, kol laikina atmintis užsipildys iki tam tikros  $n$  eilės ir iš gautos sekos yra skaičiuojamas vidurkis (vidurkis skaičiuojamas todėl, nes pagal duome-



**3.17 pav.** Kojos pakilimo ir kojos prisilietimo prie žemės ilgio variacijos kitimas slenkant langui, po transformacijos

**3.5 lentelė** Klasifikavimo rezultatas, naudojant naivų Bayes klasifikatorių

	Co	Pt
Tikslumas	0,800	0,800
Taiklumas	0,714	1,000

nų pasiskirstymą, kuris parodytas 3.15 pav., pasiskirstymų vidurkiai skiriasi, tačiau variacijos lieka tokios pačios) ir apskaičiuota reikšmė naudojama kaip naujas požymis klasifikatoriaus apmokymui, bei tikrinimui.

Dėl gaunamų duomenų matmenų ir jų mažo kiekio (kuris priklauso nuo laikinos atminties dydžio, į kurią rašomos naudos erdvės vertės), yra panaudotas paprasčiausias naivaus Bayes klasifikatoriaus mechanizmas. Duomenys nesikartojo jokiam algoritmo modulyje. Viso erdvės sudarymui panaudota 400 (nuo 1 iki 400), apmokymui 500 (nuo 501 iki 1000), testavimui 500 (1001 iki 1500) reikšmių. Laikinos atminties ilgis pasirinktas 50 verčių, iš kurių skaičiuojamas vidurkis. Apmokymo metu iš viso 10 vidurkiai kiekvienai subjektų grupei, testavimo metu iš viso 10 vidurkiai kiekvienai subjektų grupei. Gauti tikslumo ir taiklumo rezultatai pateikti 3.5 lentelėje.

Gauti klasifikavimo tikslumas yra 0,800, kas viršija prieš tai naudotų metodų tikslumą. Pirmos grupės klasifikavimo taiklumas yra 0,714, antros grupės klasifikavimo taiklumas 1,000. Detalesnė klasifikavimo mechanizmo veikimo apžvalga yra pateikta Požymių klasifikavimo programos kūrimo poskyryje.





## 4. Signalų analizės programos įgyvendinimas

Šiame skyriuje apžvelgta įgyvendintas programinis kodas, jo veikimo architektūra. Ankstesniuose poskyriuose argumentuotai pažvelgti galimi analizės metodai, požymiai ir klasifikatoriai. Visi rezultatai panaudoti projektuojant galutinį sprendimą. poskyryje 4.1. pateikta bendra programos algoritmo veikimo schema, poskyryje 4.2. pateikta algoritmo klasifikavimo veikimo schema, poskyryje 4.3. apžvelgta galutinė programa.

### 4.1. Bendro programos algoritmo schemas sudarymas

Ankstesniame skyriuje apžvelgta bendra programos veikimo schema. Bendros schemas pavyzdys yra pateiktas 3.1 pav. Šiame poskyryje pateikta detali algoritmo schema ir aptarta kiekviena jo bloko paskirtis, bei jame naudojamas metodas.

Viso programa sudaryta iš trijų modulių:

- Požymių erdvės sudarymas;
- Klasifikatoriaus apmokymas;
- Klasifikatoriaus tikrinimas.

Kiekvieno modulio pradžioje yra vykdomas pirminio signalo apdorojimas pagal schemą, pavaizduota 4.1 pav. Pirminis bloko uždavinys yra nereikalingų dažninių komponentų pašalinimas. Matlab aplinkoje signalo filtravimas yra atliekamas 9 eilės *Butterworth* filtru. Kodas pateiktas 4.1 programiniame kode. Pirmasis, aukštų dažnių filtras, skirtas pašalinti signalo nuolatinei komponentei. Antras, žemų dažnių filtras, skirtas pašalinti aukšto dažnio triukšmą, kuris neneša visiškai jokios naudingos informacijos. Abiejų filtrų eilė yra 9-ta, žemų dažnių filtro ribinis dažnis parinktas 40 Hz. Duomenys diskretizuojami 100 Hz dažniu, vadinasi didžiausias galimas signalo dažnis yra 50 Hz. Didžiausias žmogaus generuojamas dažnis ėjimo metu, remiantis šaltiniu, yra 20 Hz. Užtikrintumui parinktas 40 Hz dažnis. Nuolatinė dedamoji pašalinama su aukšto dažnio filtru, kurio ribinis dažnis yra 1 Hz. Nuolatinė dedamoji neneša jokios informacijos apie eisena, kadangi ji tiktais nurodo naudojamų jutiklių jautrumą.

Signalų skirstymas pagal žinomą fizinę veiklą Matlab aplinkoje įgyvendintas kodo fragmentas yra pateikiamas 4.2 programiniame kode. Pirmiausiai, "Pt\_t" kintamojo struktūroje yra saugoma kairės kojos Parkinsono liga sergančių subjektų duomenys. Kiekvienas signalas yra priskiriamas prie *signal* kintamojo, su kuriuo toliau yra tęsiamas apdorojimo procesas. Jeigu signalas nėra lygus nuliui, tuomet jis įdedamas į laikinąją atmintį. Taip signalas yra tikrinamas

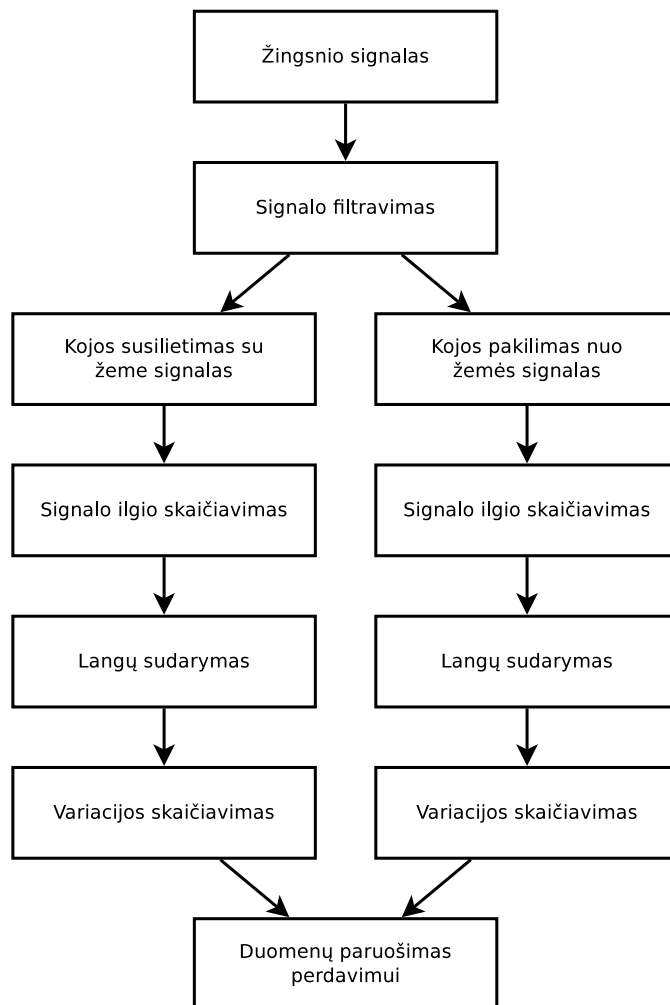
**4.1 programinis kodas.** Signalų filtravimas dviem Butterworth filtrais

```
[B,A] = butter(9, 1/50, 'high');  
[BB,AA] = butter(9, 40/50, 'low');  
output = filter(BB, AA, filter(B, A, input));
```



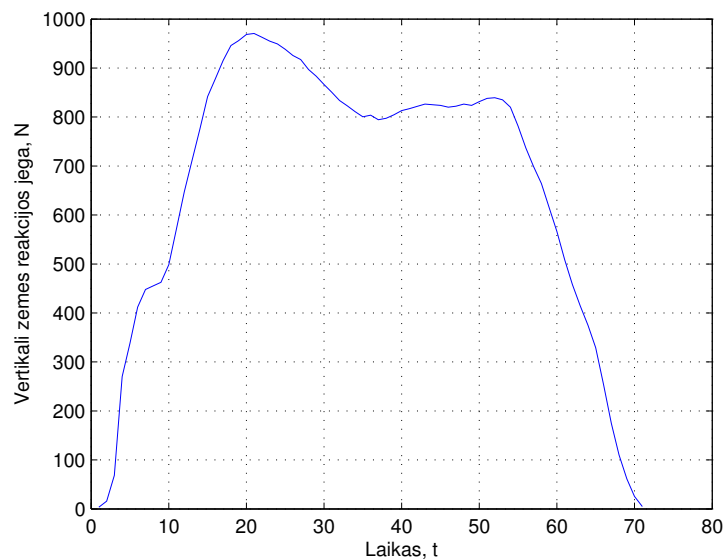
#### 4.2 programinis kodas. Kontakto su žeme signalo išskyrimo programos kodo fragmentas

```
for i=1:size( Pt_t, 2 )
    fprintf( '. ' );
    signal = Pt_t{i};
    buffer = [];
    k = 0;
    for j=1:size(signal, 1)
        if ( signal(j) ~= 0 )
            buffer = [ buffer signal(j) ]; %#ok<AGROW>
        else
            if ~isempty(buffer) && length(buffer) > 10 && length(buffer) < 200
                % Store the buffer
                data.Pt{ length(data.Pt)+1 } = buffer;
                % Clear the buffer
                if ( j - k > 50 && j - k < 200 )
                    data.Pt2( length(data.Pt2)+1 ) = j - k;
                end
                k = j; % Start of stance
                buffer = [];
            else
                buffer = [];
            end
        end
    end
end
end
```



**4.1 pav.** Pirminio signalo apdorojimo schema

iki tol, kol signalas tampa lygus nuliui ir tęsiamas tolimesnis apdorojimas. Apdorojimas susideda iš signalo ilgio patikros. Jeigu signalas yra trumpesnis už 10 verčių, arba turint omenyje, kad signalas diskretizuojamas 100 Hz, tai 0,1 s, vadinasi, signalas yra tiesiog aukšto dažnio triukšmas arba blogas pavyzdys ir signalas yra atmetamas. Jeigu signalas yra ilgesnis už 200 verčių (2 s), vadinasi, duomenų rinkimo metu įvyko klaida ir koja per tokį laiką nebuvo pakelta. Tokia klaida gali būti sukelta, kuomet subjektas ne eina, o stovi ant dviejų kojų arba tik ant kairės kojos. Jeigu visi kriterijai patenkinami, signalas yra priimamas ir kraunamas į laikinąją atmintį, pavadinimu “*data.Pt*”. Signalas, kuriuo metu koja neličia žemės, yra randamas kartu su anksčiau išnagrinėtu metu. Algoritmas patikrina kiek praėjo laiko (arba kiek verčių yra priskaičiuota) nuo paskutinio užskaityto kojos ant žemės signalo ir įrašo tą signalą į laikiną atmintį, jeigu po praeito signalo nepraėjo mažiau negu 50 verčių arba 0,5 s ir ne daugiau nei 200 verčių arba 2 s. Signalas pavaizduotas 4.2 pav. Pirminio signalo apdorojimo struktūrinė schema yra pavaizduota pav. Programos įėjime yra pateikiami pirminiai žingsnio duomenys, programa lygiagrečiai iš duomenų išskiria kojos susilietimo su žeme ir kojos pakilimo nuo žemės



4.2 pav. Susilietimo su žeme signalas

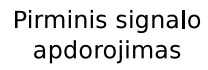
signalus, paskaičiuoja signalų ilgius, iš signalų ilgių sudaro duomenų langus, paskaičiuoja lange esančių signalų variacijos koeficientą ir paruošia gautus duomenis perdavimui.

Programos pirmo modulio struktūrinė schema pavaizduota 4.3 pav. Modulis iš pirminio signalo apdorojimo nuskaito variacijos skaičiavimus, tuomet iš esamų dviejų matmenų variacijos sudaro požymių erdvę, pritaiko LDA transformaciją su Gauso branduoliu ir gražina erdvės modelį, kuris susideda iš tikrinio vektoriaus ir duomenų vidurkio.

Programos antro modulio struktūrinė schema pavaizduota 4.4 pav. Modulis iš pirminio signalo apdorojimo nuskaito variacijos skaičiavimus. Struktūroje naudojamas erdvės modelis iš pirmojo programos modulio. Jis reikalingas tam, kad kiekvieną kartą neprojektuoti naujos erdvės iš naujo, bet projektuoti naujus duomenis į jau esamą erdvę. Tai taip pat užtikrina, kad naudojama ta pati matmenų erdvė, kas leidžia užtikrinti duomenų apibendrinimą. Skirtumas taip pirmojo ir antrojo modulio yra antro modulio išėjime yra klasifikatoriaus modelis, kuris yra apmokyto klasifikatoriaus parametrai, kurie yra panaudoti testuojant klasifikatorių.

Trečiojo, paskutinio modulio, struktūrinė schema pavaizduota 4.5 pav. Struktūroje naudojamas erdvės modelis iš pirmo modulio ir klasifikatoriaus modelis iš antro modulio. Modulio iš pirminio signalo apdorojimo nuskaito variacijos skaičiavimus. Šio modulio išėjime yra klasifikavimo rezultatas, kurį galima pateikti tiek grafiškai, tiek skaitine išraiška, išreikšta tikslumo ir taiklumo verte.

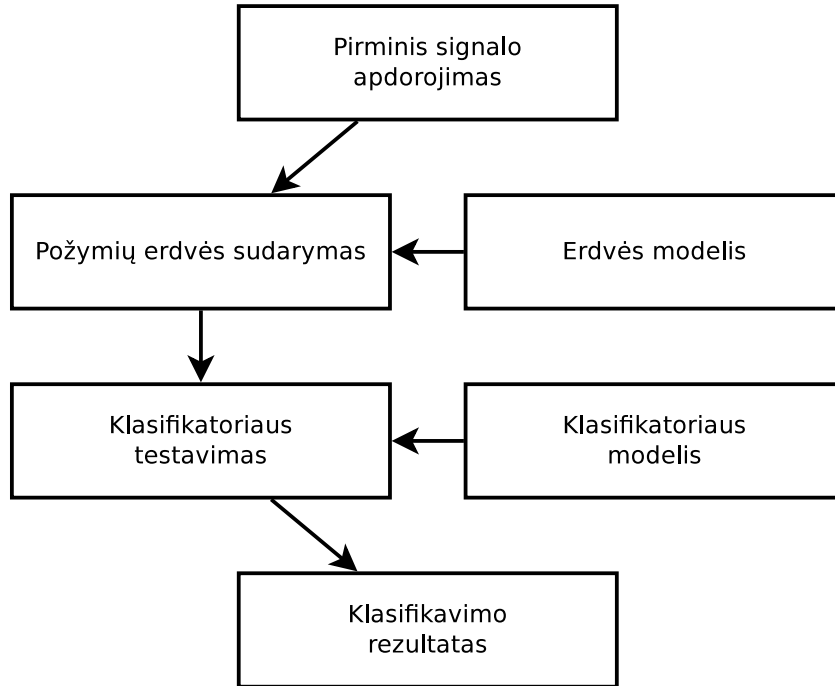
Apibendrinus bendrą programos veikimo schemą, toliau paanalizuosim patį klasifikavimo mechanizmą – naivųjį Bayes klasifikatorių.



**4.3 pav.** Programos pirmo modulio, požymių erdvės sudarymo, veikimo schema



**4.4 pav.** Antro modulio, klasifikatoriaus apmokymo, veikimo schema



4.5 pav. Trečio moduly, klasifikatoriaus tikrinimo, veikimo schema

## 4.2. Požymių klasifikavimo programos algoritmo schemas sudarymas

Šiame poskyryje apžvelgtas naudojamas klasifikatorius, trumpai aprašytas jo veikimo matematinis principas, įgyvendinimas ir struktūrinė schema.

Naivus Bayes klasifikatorius [17] yra paprastas tikimybinis klasifikatorius, kuris paremtas Bayes teorema su stipria duomenų nepriklausomybės prielaida, kuri išreiškiama:

$$P(\mathbf{X}|C) = \prod_{i=1}^n P(X_i|C), \quad (4.1)$$

kur  $\mathbf{X} = (X_1, \dots, X_n)$  yra požymių vektorius, o  $C$  yra klasės identifikatorius. Nepriklausomai nuo to, kad tokia duomenų atskyrimo prielaida yra labai naivi – praktikoje metodas veikia pakankamai gerai, ir yra naudojamas daugelį sudėtingesnių metodikų.

Bayes klasifikatorius  $h^*(x)$ , naudoja diskriminanto funkcijas atnaujinti klasės vėlesnes tikimybes, nurodant požymių vektorius:

$$f_i^*(x) = P(C = i|\mathbf{X} = x). \quad (4.2)$$

Pritaikius Bayes taisyklę nurodytai lygčiai, gaunamas rezultatas:

$$P(C = i|\mathbf{X} = x) = \frac{P(\mathbf{X} = x|C = i)P(C = i)}{P(\mathbf{X} = x)}, \quad (4.3)$$



#### 4.3 programinis kodas. Klasifikatoriaus apmokymo kodas

```
nb = NaiveBayes.fit( training_set.X', training_set.y );
```

#### 4.4 programinis kodas. Klasifikatoriaus tikrinimo kodas

```
ypred = nb.predict( testing_set.X' );
```

kur  $P(\mathbf{X} = x)$  yra identiška visoms klasėms, todėl yra ignoruojama. Iš to seka Bayes diskriminanto funkcija:

$$f_i^*(x) = P(\mathbf{X} = x|C = i)P(C = i), \quad (4.4)$$

kur  $P(\mathbf{X} = x|C = i)$  vadinama klasės priklausomybės tikimybės pasiskirstymas. Taigi, Bayes klasifikatorius:

$$h^*(x) = \operatorname{argmax}_i P(\mathbf{X} = x|C = i)P(C = i) \quad (4.5)$$

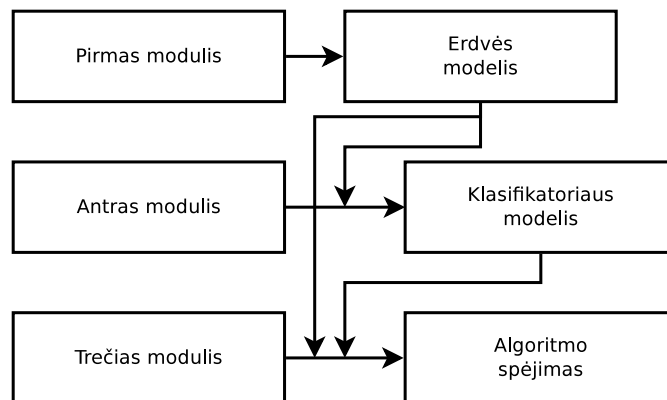
randa didžiausią vėlesnės tikimybės hipotezę, nurodžius  $x$ .

Duomenys Hilberto erdvėje yra labai gerai praskiriami tiesinė funkcija, todėl šiame sprendime yra naudojamas tokios paprastos struktūros klasifikavimo mechanizmas. Tuo labiau, kadangi duomenis yra vienmatėje erdvėje – mažinti matmenų skaičių nelieta prasmės, ieškoti klasifikavimo algoritmo, kuris geriausiai apibendrintų turimus duomenis erdvėje nėra prasmės, kadangi duomenų apibendrinimą puikiai atliko LDA su Gauso branduoliu.

Darbe panaudota Matlab aplinkoje įgyvendintas naivus Bayes klasifikatoriaus versija, kuri yra “Statistical Toolbox” paketo dalis. Klasifikatoriaus apmokymo kodas yra pateiktas 4.3 programiniame kode. Jo rezultate, kintamajam “nb” yra saugomas klasifikatoriaus modelis. Klasifikatoriaus tikrinimo kodas yra pateiktas 4.4 programiniame kode. Rezultate, kintamajam “ypred” yra saugomas programos spėjimas.

Struktūriškai klasifikatoriaus apmokymas yra pateiktas 2 priede 6.1 pav. Programa pradeda nuo signalo variacijos duomenų priėmimo iš pirminio signalo apdorojimo. Pagal ankščiau sukonstruotą erdvės modelį, projektuojami nauji signalo požymiai ir atliekamas klasifikatoriaus apmokymas.

Klasifikatoriaus tikrinimas yra vykdomas pagal 2 priede pateiktą schemą 6.2 pav. Programa pradeda tai pat, kaip ir klasifikatoriaus apmokymas iki “Erdvės sudarymas pagal požymius” struktūros bloko. Klasifikatoriaus modelis panaudojamas iš prieš tai atlikto apmokymo ir atliekamas požymių erdvės klasifikavimas. Rezultatai pateikiami neužtikrintumo matrica. Duomenys išvedami vartotojui tikslumo ir taiklumo skaitiniais įverčiais, ir grafiniu pavidalu.



4.6 pav. Galutinės programos struktūrinė schema

### 4.3. Signalų analizės programos įgyvendinimas

Šiame skyriuje apžvelgsime jau galutinę programą, aptarsime jos pritaikymo problemas, bei galimas pritaikymo platformas.

Galutinė programos schema yra pavaizduota 4.6 pav. Kaip aptarta ankstesniuose poskyriuose – programa susideda iš trijų duomenų apdorojimo modulių. Kiekvieno modulio įėjimas yra toks pats, tačiau kiekvieno modulio išėjimas yra skirtingas. Kiekvienas žemiau esantis modulis naudoja virš jos esančio modulio darbo rezultatą. Kitiškai svarbu yra kiekvienam moduliui pateikti skirtingus duomenis.

Darbe pateiktas metodas leidžia efektyviai atpažinti ar subjektas serga Parkinsono liga. Didžiausias metodo trūkumas yra tas, kad ji neatsižvelgia į kitus Parkinsono ligos simptomus – drebulys, eisenos sustingimas. Drebulys gali pasireikšti ne tik plaštakos raumenyse, tačiau ir kaklo srityje. Tai neleidžia tiksliai apibrėžti kurią kūno vietą reikia stebėti ir rinkti duomenis tyrimui. Eisenos sustingimas taip pat yra sunkiai apibrėžtas faktorius, kadangi jo aptikimas yra labai didelis iššūkis signalų apdorojimo srityje. Geriausią ką šiuo metu gali pasiūlyti mokslas, stebint tokius ligos simptomus – paciento stebėjimas kameros pagalba, jo veiklos automatinis nustatymas. Drebulys subjektui dažniausiai pasireiškia, kai jo kūnas yra visiškai atsipalaidavęs, t.y. kai subjektas stovi, sėdi, guli, kai jis išlaiko statišką poziciją, tuomet pasireiškia drebulys. Kameros pagalba galima nustatyti kokioje pozicijoje yra subjektas, tačiau identifikuoti drebulį yra labai sudėtinga, jeigu naudojama kamera yra mažos rezoliucijos. Drebulį veiklos nustatymo algoritmas gali palaikyti tiesiog pašaliniu triukšmu, kaip šešėlio sudarymą ant stebimo paviršiaus.

Darbas ties Parkinsono ligos identifikavimu ir jo diagnozavimui reikia įdėti dar daug darbo, tačiau besivystant kompiuteriniai technikai, bei atsirandant vis naujiems algoritmams nestandartinėms problemoms spręsti – šansas, kad ateityje šios ligos diagnozavimas pagerės, išlieka labai didelis.



## 5. Signalų analizės programos patikra

Šiame skyriuje parengtas ir įgyvendintas algoritmo patikros planas. Programos patikra yra kritinis aspektas jos patikimumo tikrinimui. Algoritmą galima tikrinti mažinant apmokymo verčių skaičių ir didinant tikrinimo verčių kiekį – taip sužinota kiek mažiausiai verčių reikia metodui, norint pilnai apibendrinti turimus duomenis.

Tikrinant kiekvieno žmogaus žingsnio signalus, paaiškėjo, kad skirtingi subjektai sugeneruoja kitokį tikrinimą atliktų žingsnių skaičių, todėl tikrinimas atliekamas ne atskiriant konkrečius žmones, tačiau jau išskirtus kojos pakilimo ir kojos prisilietimo prie žemės signalus, konkrečiau – iš tų signalų suformuotus jų ilgių duomenų langus.

Po pirminio duomenų apdorojimo, kontrolinių subjektų grupėje liko 1790 duomenų langų, Parkinsono subjektų grupėje liko 1619 duomenų langų, todėl nuspręsta iš kiekvienos grupės pasiimti po 1500 langų duomenų, visi jie padalinti po lygias tris dalis: viena dalis erdvės sudarymui, antra dalis klasifikatoriaus apmokymui, trečia dalis klasifikatoriaus testavimui. Jokie duomenys jokioje dalyje pasikartotinai nesikartoja. Tai yra kritinis faktorius mašiniam apmokyme [9]. Taip pat reikia užtikrinti duomenų praskyrimo ir klasifikatoriaus apmokymui dešimtį kartų didesnių duomenų skaičių, negu yra tų pačių duomenų matmenų. Kadangi prieš duomenų praskyrimą turima tik  $d = 2$  matmenys, tai teoriškai užtektų ir  $n = 21$  ( $n/d > 10$ ), tačiau matmenų praskyrimas vyksta su 500 taškų, kas tik užtikrina apibendrinimą.

### 5.1. Eksperimentų plano rengimas

Šiame poskyryje aptarti galimi eksperimentiniai algoritmo patikros planai, nurodytas galimos duomenų pateikimo metodikos.

Pirminį patikros planą sudaro pateikiamų duomenų skaičiaus mažinimas į kiekvieno algoritmo žingsnį: matmenų erdvės sudarymas, klasifikatoriaus apmokymas, klasifikatoriaus testavimas. Duomenis galima mažinti tiesiniu būdu – matmenų mažinime naudoti pirmus 400 duomenų, 100 praleisti, nuo 501 iki 1000 paduoti klasifikatoriaus apmokymui ir nuo 1001 iki 1500 pateikti klasifikatoriaus tikrinimui. Taip pat duomenis galima pateikti kas kelintą  $n$  narį – kadangi iš viso yra trys algoritmo žingsniai, tuomet  $n = 3$ . Taip į matmenų mažinimo algoritmą pateikiamas 1 rinkinys, 4 rinkinys, 7 rinkinys. Klasifikatoriaus apmokymui pateikiamas 2 rinkinys, 5 rinkinys, 8 rinkinys. Klasifikatoriaus tikrinimui pateikiamas 3 rinkinys, 6 rinkinys, 8 rinkinys. Toks duomenų pateikimas leidžia pasiekti “apibendrinto” metodo.

Taip pat, kadangi sprendime naudojamas laikinos atminties modulis, kuris reikalingas saugoti pirmąjį diskriminantą kiekvienos variacijos poros, eksperimentiniu būdu galima spręsti koks lango dydis nurodytai užduočiai atlikti geriausiai tinka. Atliekant tokį tikrinimą svarbu palaikyti kitų parametrų vienodumą, todėl į kiekvieną iš trijų modulių pateikiamas pastovus duomenų skaičius, kuomet keičiamas laikinos atminties dydis. Tikrinimas taip pat nuspręstas kiek minimaliai iš paciento eisenos turi būti išskirta požymių grupių, kad sistema gražintų geriausią klasifikavimo tikslumo ir taiklumo rezultata.

**5.1 lentelė** Klasifikavimo tikslumo ir taiklumo rezultatai, tiesiškai mažinant pateikiamų duomenų matmenų mažinimo algoritmui

		Taškų rinkinio skaičius					
		500	400	300	200	100	21
Co	Tikslumas	0,700	<b>0,800</b>	0,450	0,450	0,500	0,450
	Taiklumas	0,750	<b>0,875</b>	0,467	0,429	0,500	0,400
Pt	Tikslumas	0,700	<b>0,800</b>	0,450	0,450	0,500	0,450
	Taiklumas	0,667	<b>0,750</b>	0,400	0,462	0,500	0,467

**5.2 lentelė** Klasifikavimo tikslumo ir taiklumo rezultatai, tiesiškai mažinant pateikiamų duomenų klasifikatoriaus apmokymo algoritmui

		Taškų rinkinio skaičius					
		500	400	300	200	100	21
Co	Tikslumas	<b>0,800</b>	<b>0,800</b>	0,750	0,750	0,650	—
	Taiklumas	<b>0,875</b>	<b>0,875</b>	0,857	1,000	1,000	—
Pt	Tikslumas	<b>0,800</b>	<b>0,800</b>	0,750	0,750	0,650	—
	Taiklumas	<b>0,750</b>	<b>0,750</b>	0,692	0,667	0,588	—

## 5.2. Duomenų eksperimentams rengimas

Šiame poskyryje aptartas duomenų eksperimentams rengimas, aprašytas planuojamas duomenų kiekis, naudojamas tikrinimo metu.

Pradžioje yra tikrinamas matmenų mažinimo algoritmas, mažinant pateikiamų duomenų skaičių. Struktūriškai nuspręsta, kad kiekvienas modulis turi lygiai po 500 rinkinių. Tikrinimas pradedamas nuo 500 rinkinių ir mažinamas kas 100 rinkinių. Vadinasi, iš viso matmenų mažinimo algoritmui pateikiama 500, 400, 300, 200, 100 ir minimalus 21 rinkinių skaičius. Į apmokymo lygmenį pateikiamas pastovus duomenų skaičius – 500 (nuo 501 iki 1000) rinkinių. Laikinos atminties dydis palaikomas pastovus – 50 verčių. Po matmenų mažinimo žingsnio tikrinimo, rinkinių skaičius tokia pat metodika mažinamas klasifikatoriaus apmokymui, matmenų mažinimo metodui pateikiant pastovų rinkinių skaičių. Šioje stadijoje duomenų skaičius nėra mažinamas iki 21 rinkinio skaičiaus, kadangi neformuojamas naudojamas lango ilgio, 50, duomenų skaičius. Klasifikatoriaus testavimo atveju duomenų dydis nėra mažinamas.

Laikinos atminties ilgio tikrinimo metu panaudoti tokie laikinos atminties ilgiai: 10, 30, 50, 70, 100, 130 ir 150. Kiekvienas toks tikinimas atliekamas su duomenų kiekiu, kuris parodė geriausią klasifikavimo tikslumo ir taiklumo rezultatą, aprašytą ankstesniame paragrafe. Panaudoti rezultatai atitinkamai užtikrins optimalų duomenų kiekio pasiskirstymą, geriausiam klasifikavimo tikslumo ir taiklumo rezultatui pasiekti.

## 5.3. Programos patikros rezultatai

Šiame poskyryje pateikti ir aptarti patikros tikslumo ir taiklumo rezultatai, išanalizuoti galimi sistemos parametrai, kurie gali būti derinami: duomenų skaičius, kuris naudojamas



**5.3 lentelė** Klasifikavimo tikslumo ir taiklumo rezultatai, netiesiškai mažinant pateikiamų duomenų matmenų mažinimo algoritmui

		Taškų rinkinio skaičius					
		500	400	300	200	100	21
Co	Tikslumas	0,550	0,450	0,600	0,600	0,450	<b>0,650</b>
	Taiklumas	0,533	0,467	0,583	0,583	0,400	<b>0,667</b>
Pt	Tikslumas	0,550	0,450	0,600	0,600	0,450	<b>0,650</b>
	Taiklumas	0,600	0,400	0,625	0,625	0,467	<b>0,636</b>

**5.4 lentelė** Klasifikavimo tikslumo ir taiklumo rezultatai, netiesiškai mažinant pateikiamų duomenų klasifikatoriaus apmokymo algoritmui

		Taškų rinkinio skaičius					
		500	400	300	200	100	21
Co	Tikslumas	0,550	0,450	0,500	0,400	<b>0,550</b>	–
	Taiklumas	0,553	0,455	0,500	0,400	<b>1,000</b>	–
Pt	Tikslumas	0,550	0,450	0,500	0,400	<b>0,550</b>	–
	Taiklumas	0,600	0,444	0,500	0,400	<b>0,526</b>	–

kiekvienam programos moduliui, bei atminties dydis, kuris naudojamas pirmajai diskriminanto vertei saugoti.

Pirma patikra atlikta tiesiškai mažinant duomenų kiekį, kuris yra paduodamas matmenų mažinimo algoritmui, taip siekiant nustatyti koks duomenų kiekis geriausiai tinka naujai požymių erdvei projektuoti. Eksperimentiniai duomenys yra pateikti 5.1 lentelėje. Vertinant matmenų sudarymą pagal klasifikatoriaus duodamą rezultatą, geriausias duomenų kiekis, skirtas naujai matmenų erdvei sudaryti yra 400 erdvės taškų. Tokiu atveju pasiekiamas 0,875 Pt ir 0,750 Co taiklumas.

Antra patikra atlikta tiesiškai mažinant duomenų kiekį, kuris yra paduodamas klasifikatoriaus apmokymui, taip siekiant nustatyti, kokio duomenų kiekio reikia klasifikatoriui, kad jis sugebėtų apibendrinti duomenis, naudojamus testavimo stadijoje. Eksperimentiniai duomenys yra pateikti 5.2 lentelėje. Šio tikrinimo metu paaiškėjo, kad geriausiai klasifikatoriaus apmokymui tinka du duomenų rinkiniai – po 500 ir po 400. Tokiu atveju pasirenkamas didesnis duomenų kiekis, taip užtikrinant bendrų duomenų apibrėžtumą.

**5.5 lentelė** Klasifikavimo tikslumo ir taiklumo rezultatai, mažinant laikinosios atminties dydį

		Laikinosios atminties dydis						
		150	130	100	70	50	30	10
Co	Tikslumas	0,500	0,500	0,700	0,500	<b>0,800</b>	0,500	0,570
	Taiklumas	0,500	0,500	0,750	0,500	<b>0,875</b>	0,500	0,585
Pt	Tikslumas	0,500	0,500	0,700	0,500	<b>0,800</b>	0,500	0,570
	Taiklumas	0,500	0,500	0,667	0,500	<b>0,750</b>	0,500	0,559



## 6. Rezultatų apibendrinimas

Darbo metu ištirti galimi žingsnio požymiai, kuriais remiantis galima sėkmingai atpažinti Parkinsono ligą pagal subjekto eiseną. Nustatyta, kad dažninės žingsnio komponentės, koreliacijos koeficientas, dviejų maksimumų ir vieno minimumo požymiai neturi pakankamai informacijos Parkinsono ligos atpažinimui. Daugiausiai informacijos turi kojos prisilietimo prie žemės ir kojos pakilimo nuo žemės signalo laiko variacijos požymis.

Turint duomenis, kurie turi daugiausiai informacijos ligos identifikavimui, toliau patikrinta galimų matmenų praskyrimo metodų pritaikymas. Tiesiniai PCA ir LDA transformacijos metodai požymių erdvės tinkamai nepraskyrė. Geriausiai užduotį LDA transformacija su Gauso branduoliu. Pritaikius nagrinėjamas transformacijas duomenų grupėms, pastebėta kaip branduolio metodo pritaikymas gali padidinti matmenų mažinimo algoritmo efektyvumą.

Turimus vienmačius duomenis realiu laiku geriausiai klasifikavo naivus Bayes klasifikatorius, tačiau klasifikavimo tikslumas ir taiklumas yra nepatenkinamas, tikslumas siekė tik 50,8 %. Pritaikius papildomą metodikos žingsnį – laikinosios atminties bloką, kuriame saugomas pirmas transformacijos diskriminantas, klasifikavimo tikslumas pagerintas iki 80 %. Rezultatas nurodo kas penkto diagnozavimo rezultato klaidingumą. Turint omenyje, kad klinikinės diagnostikos sprendimų taiklumas yra nuo 74 % iki 90 % [1], sistemos darbo rezultatas kitų produktų palyginime atrodo patenkinamai.

Pagrindinis darbo uždavinys buvo sukurti sistemą, kuri gebėtų atpažinti Parkinsono ligą pagal galimus ligos požymius, nagrinėjant subjektus pagal jų eisenos ypatybes. Toks uždavinys yra įvykdytas, tačiau egzistuoja 1/5 dalies netikslumas. Toks netikslumas argumentuojamas kiekvieno žmogaus eisenos unikaliomis savybėmis. Tokia neigiama metodo savybė yra pašalinama, atliekant diagnozę mažiausiai 5 kartus.

Darbe planuotas naudoti dirbtinių neuronų tinklas panaudotas nebuvo. Duomenys požymių erdvėje buvo lengvai atskiriami tiesine funkcija, todėl naudoti kompleksinio dirbtinių neuronų tinklų klasifikatorių nėra prasmės. Panaudotas paprastesnis naivus Bayes klasifikatorius.

Naudojamų subjektų skaičius klasifikatoriaus apmokymui ir tikrinimui panaudotas nebuvo. Iš viso, klasifikatoriaus apmokymui planuota panaudoti 60 Parkinsono liga sergančių ir 50 kontrolinių subjektų. Klasifikatoriaus tikrinimui planuota panaudoti 33 sergančių ir 23 sveikų subjektų. Kiekvienas iš subjektų generuoja skirtingą skaičių patikrintų žingsnių signalų, iš kurių toliau skaičiuojami požymiai. Duomenų analizės metu, norint suvienodinti ir tuo pačiu supaprastinti uždavinio nagrinėjimą, pirmiausiai iš kiekvieno subjekto išskirti žingsnio fazės signalai ir jie visi sujungti į vieną matricą. Tuomet egzistuoja dvi matricos, kurios priklauso skirtingai subjektų grupei. Toliau dalinti duomenis pagal kiekvieno subjekto sugeneruotų duomenų kiekį nėra prasmės, kadangi pirmiausiai yra naudojamas skirtingas subjektų skaičius kiekvienos grupės atžvilgiu ir yra garantuojamas duomenų skaičiaus neatitikimas. Skaičiaus atitikimas yra būtinas erdvės transformacijai atlikti, todėl reikia suvienodinti esamų duomenų

skaičių. Tikrinimo atveju turi būti atliekama tokia pati procedūra. Dėl šios priežasties buvo parinktas visų esamų duomenų dalinimas į tris lygias dalis kiekvieno programos modulio įgyvendinimui.

Sistema buvo planuojama dirbti nerealiu laiku. Toks uždavinys buvo įvykdytas. Darbe toks pasirinkimas argumentuojamas pačio eisenos variacijos chaotišku kitimu. Reikalingas ilgas, 50 ilgio vertės variacijos verčių vektorius, norint efektyviai nustatyti ar yra ligos požymiai ir galima teigti, kad subjektas turi Parkinsono ligą.

Gautus rezultatus galima pagerinti, išnagrinėjus daugiau kontrolinių subjektų eisenos ypatybių, bei Parkinsono liga sergančių subjektų eisenos ypatybes. Iš viso buvo išnagrinėti 93 sergantys subjektai ir 73 sveikas subjektas. Iš sveikų subjektų iš viso buvo išskirta 3543 kojos pakilimo nuo žemės signalų, 3583 kojos prisilietimo prie žemės signalai. Iš sergančių subjektų iš viso buvo išskirta 3241 kojos pakilimo nuo žemės signalų, 3217 kojos prisilietimo prie žemės signalai. Iš gautų signalų buvo paskaičiuota jų ilgiai ir naudojantis slankiojančio lango metodu – jų variacija. Iš viso, sveikų subjektų buvo požymių buvo 1790, sergančių subjektų požymių 1619 verčių. Nurodyto darbo rezultatus gali patikslinti tik dar didesnis subjektų skaičius, bei ilgesnis eisenos laikas, kuris šių duomenų atveju buvo tik 2 min.

Darbe panaudotos priemonės lūkesčius pateisino dalinai. Išskirtas požymis identifikuoja ligos požymius tik esant ilgos eisenos prielaidai – subjektas turi atlikti eisenos patikrą ilgiau negu 2 min. Rekomenduojama eisenos trukmė yra 5 min. Tokiu atveju sistema geriausiai identifikuos ligos simptomą pagal parinktus požymius. Taip pat klasifikavimo rezultatus gali pagerinti abiejų kojų naudojimas požymių išskyrimo metu, kadangi eisenos nesimetriškumas gali galioti ne tik kairiai kūno pusei, tačiau ir dešinei.



## Literatūros ir informacinių šaltinių sąrašas

- [1] Artūras; Griškevičius Julius [Griskevicius Julius ]. Aubin, Patric M.; Serackis. A ground reaction force artificial neural network classifier for the diagnosis of parkinson's disease. *Proceedings of XXIIIrd ISB2011 Congress*, pages 428–437, July 2011.
- [2] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
- [3] Chih-Chung Chang and Chih-Jen Lin. Libsvm – a library for support vector machines.
- [4] Arnaud Delval, Julia Salleron, Jean-Louis Bourriez, Séverine Bleuse, Caroline Moreau, Pierre Krystkowiak, Luc Defebvre, Patrick Devos, and Alain Duhamel. Kinematic angular parameters in pd: reliability of joint angle curves and comparison with healthy subjects. *Gait & Posture*, 28(3):495–501, 2008.
- [5] Silvi Frenkel-Toledo, Nir Giladi, Chava Peretz, Talia Herman, Leor Gruendlinger, and Jeffrey Hausdorff. Effect of gait speed on gait rhythmicity in parkinson's disease: variability of stride time and swing time respond differently. *Journal of NeuroEngineering and Rehabilitation*, 2(1):23, 2005.
- [6] Jeffrey M. Hausdorff, Merit E. Cudkowicz, Renée Firtion, Jeanne Y. Wei, and Ary L. Goldberger. Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in parkinson's disease and huntington's disease. *Movement Disorders*, 13(3):428–437, 1998.
- [7] Jeffrey M. Hausdorff, Susan L. Mitchell, Renée Firtion, C. K. Peng, Merit E. Cudkowicz, Jeanne Y. Wei, and Ary L. Goldberger. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and huntington's disease. *Journal of Applied Physiology*, 82(1):262–269, 1997.
- [8] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *ArXiv Mathematics e-prints*, January 2007.
- [9] A.K. Jain, R.P.W. Duin, and Jianchang Mao. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, jan 2000.
- [10] 1998 Kevin Murphy. Hidden markov model (hmm) toolbox for matlab.
- [11] E. KNUTSSON. An analysis of parkinsonian gait. *Brain*, 95(3):475–486, 1972.
- [12] H.H. Manap, N.M. Tahir, and A.I.M. Yassin. Statistical analysis of parkinson disease gait classification using artificial neural network. In *Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on*, pages 060–065, dec. 2011.



- [13] Vojtech Franc Michal. I. Schlesinger, Vaclav Hlavac. Statistical pattern recognition toolbox, September 2000.
- [14] G.B. Moody, R.G. Mark, and A.L. Goldberger. Physionet: a web-based resource for the study of physiologic signals. *Engineering in Medicine and Biology Magazine, IEEE*, 20(3):70–75, may-june 2001.
- [15] Maksim Norkin. Human activity recognition using Hidden Markov Models, September 2011.
- [16] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, feb 1989.
- [17] Irina Rish. An empirical study of the naive bayes classifier. In *In proceeding of IJCAI-01 workshop on Empirical Methods in AI*, pages 41–46, November 2001.
- [18] Jonathon Shlens. A tutorial on principal component analysis. <http://www.sn1.salk.edu/~shlens/pub/notes/pca.pdf>, December 2005.
- [19] J. Stamatakis, J. Cremers, D. Maquet, B. Macq, and G. Garraux. Gait feature extraction in parkinson’s disease using low-cost accelerometers. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 7900–7903, 30 2011-sept. 3 2011.
- [20] B. W. Stansfield, S. J. Hillman, M. E. Hazlewood, and J. E. Robb. *Gait & posture*, volume 23, chapter Regression analysis of gait parameters with speed in normal children walking at self-selected speeds, pages 288–294. Elsevier Science, Apr 2006.
- [21] I. Tien, S.D. Glaser, and M.J. Aminoff. Characterization of gait abnormalities in parkinson’s disease using a wireless inertial sensor system. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3353–3356, 31 2010-sept. 4 2010.
- [22] M. Welling. Fisher linear discriminant analysis. *Department of Computer Science, University of Toronto*, 2005.
- [23] Wikipedia. Filter (signal processing), 05 2012. [http://en.wikipedia.org/wiki/Filter\\_\(signal\\_processing\)](http://en.wikipedia.org/wiki/Filter_(signal_processing)).
- [24] Wikipedia. Parkinson’s disease, 04 2012. <http://en.wikipedia.org/wiki/Parkinsons>.
- [25] Yunfeng Wu and S. Krishnan. Statistical analysis of gait rhythm in patients with parkinson’s disease. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(2):150–158, april 2010.





## Santrauka

The goal of the Bachelor thesis was to develop an application, which is capable of Parkinson's disease recognition from gait analysis. The work was started from the alternatives methods review, with discussion of drawbacks and advantages. Based on this review, the methods, which will be used in this work, was derived. The second step was to find the features of the signals, which would most effectively separate the Parkinson's subjects from control subjects. The most logical feature had problems in feature space, so the dimensional reduction method had to be applied. The classification mechanism was chosen very simple, because of the one-dimensional data in feature space and classification only of two classes with no further information. System verification confirmed, that system is able to recognize Parkinson's subjects, but additional experiments with more data must be made.



## PRIEDAI

### 1 priedas. PCA įgyvendinimas, panaudojus skirtingas metodikas.

#### 6.1 programinis kodas. PCA įgyvendinimas, panaudojus SVD

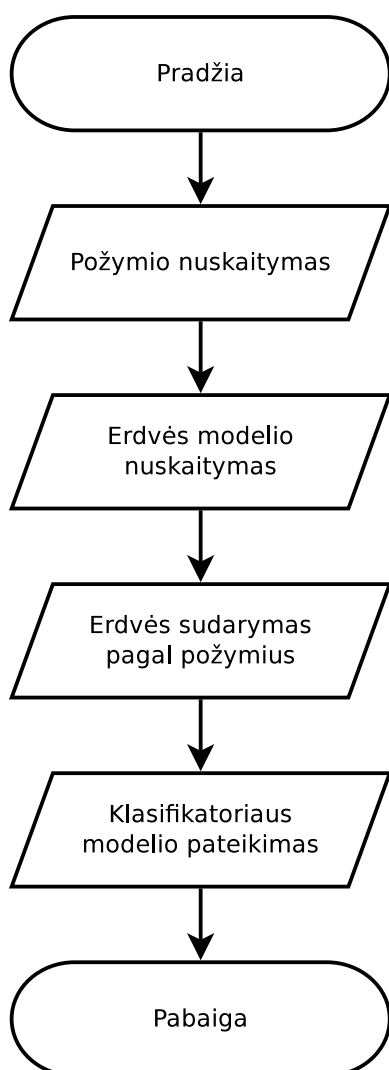
```
function [signals,PC,V] = pca2(data)
% PCA2: Perform PCA using SVD.
% data – MxN matrix of input data
% (M dimensions, N trials)
% signals – MxN matrix of projected data
% PC – each column is a PC
% V – Mx1 matrix of variances
[M,N] = size(data);
% subtract off the mean for each dimension
mm = mean(data,2);
data = data – repmat(mm,1,N);
% construct the matrix Y
Y = data' / sqrt(N–1);
% SVD does it all
[u,S,PC] = svd(Y);
% calculate the variances
S = diag(S);
V = S .* S;
% project the original data
signals = PC' * data;
```



## 6.2 programinis kodas. PCA įgyvendinimas, panaudojus tikrinių vektorių dekompoziciją

```
function [signals,PC,V] = pca1(data)
% PCA1: Perform PCA using covariance.
% data – MxN matrix of input data
% (M dimensions, N trials)
% signals – MxN matrix of projected data
% PC – each column is a PC
% V – Mx1 matrix of variances
[M,N] = size(data);
% subtract off the mean for each dimension
mn = mean(data,2);
data = data – repmat(mn,1,N);
% calculate the covariance matrix
covariance = 1 / (N–1) * data * data';
% find the eigenvectors and eigenvalues
[PC, V] = eig(covariance);
% extract diagonal of matrix as vector
V = diag(V);
% sort the variances in decreasing order
[junk, rindices] = sort(–1*V);
V = V(rindices);
PC = PC(:,rindices);
% project the original data set
signals = PC' * data;
```

## 2 priedas. Požymio klasifikavimo programos algoritmo schemas



**6.1 pav.** Klasifikatoriaus apmokymo schema

