



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
KOMPIUTERIJOS KATEDRA

Magistro baigiamasis darbas

Rekurentinių diagramų tipo metodai elektroencefalogramų analizei

Atliko:

Marekas Molis

parašas

Vadovas:

Prof. dr. Tadas Meškauskas

Vilnius
2018

Turinys

Santrauka	4
Summary	5
Įvadas	6
1. Rekurentinės diagramos	8
1.1. Rekurentinių diagramų apibrėžimas	8
1.2. Signalų būsenų sąvoka	8
1.3. Atstumo tarp dviejų būsenų skaičiavimas	9
1.4. Parametrų pasirinkimas	10
1.4.1. Būsenų dimensijos pasirinkimas	11
1.4.2. Vėlinimo parametro pasirinkimas	12
1.4.3. Slenkstinio atstumo pasirinkimas	13
2. Rekurentinių diagramų kiekybinė analizė	14
2.1. Rodikliai	14
2.1.1. Rodikliai besiremiantys juodų taškų tankiu	14
2.1.2. Rodikliai besiremiantys įstrižomis linijomis	15
2.1.3. Rodikliai besiremiantys vertikaliomis linijomis	15
3. Elektroencefalogramos	17
3.1. EEG signalų gavimas	17
3.2. 10-20 sistema	17
3.3. EEG taikymas diagnostikoje	19
3.4. EEG artefaktai	19
4. Atraminių vektorių klasifikatorius	20
4.1. Parametrų pasirinkimas	21
5. Dirbtiniai neuroniniai tinklai	23
5.1. Perceptronas	23
5.2. Aktyvacijos funkcijos	24
5.3. Tinklo modelis	25
5.4. Modelio apmokymas	25
5.4.1. Neuroninio tinklo parametrų parinkimas	26
6. Atlikti tyrimai	28
6.1. Programavimo kalbos pasirinkimas	28
6.2. Naudotos bibliotekos	28
6.2.1. EEGTools	28
6.2.2. PyRQA	28
6.2.3. Scikit-learn	28
6.2.4. TensorFlow	28
6.2.5. Keras	29
6.2.6. Pandas	29

6.3.	Rekurentinių diagramų kiekybinės analizės taikymas ligų klasifikavimui	29
6.3.1.	Duomenys	29
6.3.2.	50 Hz dažnio artefakto pašalinimas iš EEG signalo	29
6.4.	Tyrimas su 16 s ilgio EEG signalų fragmentais	31
6.4.1.	Statistinė RD kiekybinių rodiklių analizė	31
6.4.2.	Klasifikavimo strategija	33
6.4.3.	Atraminų vektorių klasifikatoriaus taikymas	33
6.4.4.	Neuroninio tinklo taikymas	35
6.5.	Tyrimai su kitais EEG fragmentų ilgiais	35
6.5.1.	Duomenys ir tyrimo metodika	35
6.5.2.	Statistinė analizė	36
6.5.3.	Klasifikatorių taikymas	37
6.6.	Tyrimas su kanalais, kuriuose nėra pikų	37
6.7.	Tyrimas su pikų sekomis	39
6.7.1.	Duomenys ir tyrimo metodika	40
6.7.2.	Klasifikavimo rezultatai	40
	Išvados ir rekomendacijos	41
	Ateities tyrimų gairės	42
	Padėka	43
	Literatūros šaltiniai	44
	Priedai	47
A.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 8 sekundžių EEG fragmentams	48
B.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 16 sekundžių EEG fragmentams	49
C.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 32 sekundžių EEG fragmentams	50
D.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės EEG fragmentams, kuriuose nėra pikų	51
E.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 8 sekundžių ilgio testiniams duomenims	52
F.	Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 32 sekundžių ilgio testiniams duomenims	53

Santrauka

Rekurentinės diagramos yra plačiai naudojamos atliekant sudėtingų dinaminių sistemų analizę. Sudėtingų sistemų pavyzdžiu galėtų būti medicina, nes pastarojoje sutinkami signalai paprastai yra netiesiniai ir chaotiniai. Sudėtingų signalų klasifikavime plačiai yra naudojami rekurentinių diagramų išvestiniai rodikliai. Šiame darbe rekurentinių diagramų išvestiniai rodikliai naudojami klasifikuoti pacientus, sergančius Rolando epilepsija ir turinčius fizinius smegenų pažeidimus.

Rekurentinių diagramų kiekybinės analizės rodikliai buvo apskaičiuoti Santaros klinikų vaikų ligoninės duomenims. Buvo atlikta analizė su 8, 16, 32 sekundžių ilgio elektroencefalogramų (EEG) fragmentais. Atlikta statistinė analizė parodė skirtumus tarp dviejų pacientų grupių. Geriausias klasifikavimo rezultatas su apmokyme nedalyvavusių pacientų duomenimis buvo pasiektas su 16 s ilgio fragmentais. Atraminų vektorių klasifikatorius parodė 72% tikslumą, neuroninis tinklas - 67%.

Gydytojai darbe nagrinėjamas diagnozes atskiria nagrinėdami pacientų EEG esančius pikus. Šiame darbe buvo bandoma klasifikuoti diagnozes apskaičiavus rekurentinių diagramų kiekybinius rodiklius įvairaus ilgio pikų sekoms. Pikai buvo iškerpami iš pacientų EEG, juos aptikus specialiu automatizuotų pikų paieškos algoritmu. Priklausomybės tarp sekoje esančio pikų kiekio ir klasifikavimo tikslumo nustatyti nepavyko.

Summary

Methods of Recurrence Plot Type for Analysis of Electroencephalograms

Recurrence plots are widely used in the analysis of complex dynamic systems. A good example of complex systems could be found in medicine where the signals encountered in the latter are usually nonlinear and chaotic. Variables of recurrence quantification analysis are widely used in the classification of complex signals. In this work recurrence quantification analysis is used to classify patients with rolandic epilepsy or structural focal epilepsy.

Data for analysis was provided by Children's Hospital, Affiliate of Vilnius University Hospital Santaros Klinikos. There was an analysis of 8, 16, 32-sec-long electroencephalogram (EEG) fragments performed in this work. The statistical analysis showed differences between the two patient groups. The best classification result was achieved with EEG fragments of 16 s. SVM classifier showed an accuracy of 72%, neural network - 67%.

Doctors distinguish two types of epilepsy by analyzing epileptiform discharges also known as spikes. In this work, the author tried to classify the epilepsy types by performing the quantitative analysis of spikes sequences. Spikes were detected in patient's EEG with a special algorithm and then spikes were gathered to one sequence. The dependency between the number of spikes in the sequence and the accuracy of the classification was not determined.

Ivydas

Sudėtingų netiesinių dinaminų sistemų analizė yra netrivialus uždavinys. Rekurencija arba pasikartojamumas yra viena iš fundamentalių netiesinių dinaminų sistemų savybių [15]. Pasikartojimų aptikimas gali suteikti daug informacijos apie nagrinėjamą sistemą. Rekurentinė diagrama (RD) (angl. *recurrence plot*) yra galingas įrankis, kuris leidžia tirti periodiškumą sudėtinguose signaluose. Šis įrankis leidžia vizualizuoti signalą ir jame išvelgti tam tikras struktūras. Tačiau toks tyrimo būdas ne visada yra pakankamai formalus, todėl buvo sukurti rodikliai, kurie leidžia kiekybiškai įvertinti rekurentines diagramas. Šie rodikliai padeda klasifikuoti signalus.

Elektroencefalogramos (EEG) yra vienas iš pagrindinių įrankių centrinės nervų sistemos ligų diagnostikoje ir gydyme. Pasak neurofiziologų EEG signalai pasižymi dideliu netiesiškumu ir kompleksišku [31]. Didelės EEG signalų apimtys bei jų sudėtingumas žymiai prailgina signalų analizę. 1 valandos EEG analizė gali užtrukti 5-10 valandų [19], todėl automatinių analizės įrankių poreikis yra akivaizdus.

Rekurentinių diagramų kiekybinė analizė plačiai taikoma epilepsijos diagnostikoje, tačiau pagrindinis dėmesys yra skiriamas būsenos tarp priepuolių (angl. *interictal*) ir priepuolio būsenos (angl. *ictal*) atskirumui [9], [10]. Šie metodai taip pat naudojami sveikų žmonių ir epilepsija sergančių pacientų klasifikavimui [7], [21]. Šiame darbe rekurentinių diagramų kiekybinė analizė bus taikoma dviejų epilepsijos tipų pacientų klasifikavimui.

Epilepsija sergančių pacientų EEG pasižymi pikais. Įvairių epilepsijos tipų pikų morfologija yra panaši, todėl nustatyti tikslią diagnozę rankiniu būdu yra sudėtinga arba neįmanoma [20]. Šiame darbe yra nagrinėjamos dviejų pacientų grupių EEG: epilepsijos sukeltos dėl įvairių fizinių smegenų pažeidimų (pvz. cerebrinio paralyžiaus, displazijos, gliozės ir pan.) ir Rolando epilepsijos. Įrankis automatiškai klasifikuojantis šių dviejų grupių pacientus ženkliai palengvintų gydytojų darbą.

Šio **tyrimo tikslas** yra klasifikuoti pacientų EEG pagal diagnozę (Rolando epilepsija ar smegenų pažeidimai) naudojant RD išvestinius metodus (rekurentinių diagramų kiekybinė analizė (angl. *recurrence quantification analysis, RQA*)) ir įvertinti klasifikavimo tikslumą.

Šio darbo uždaviniai yra šie:

- Išnagrinėti dirbtinių neuroninių tinklų teoriją.
- Parašyti programą pikams iš EEG signalo iškirpti, kai yra žinomos tikslios pikų vietos EEG signale.
- Apskaičiuoti RD kiekybinius rodiklius 8, 16, 32 s EEG ilgio fragmentams, dviejų epilepsijos tipų pacientams.
- Apskaičiuoti RD kiekybinius rodiklius įvairaus ilgio pikų sekai viename signale, dviejų epilepsijos tipų pacientams.
- Atlikti gautų rodiklių statistinę analizę.
- Apmokyti atraminių vektorių klasifikatorių bei neuroninį tinklą su dviejų tipų epilepsija sergančių pacientų RD kiekybiniais rodikliais ir įvertinti klasifikavimo tikslumą su apmokyme nedalyvavusių pacientų duomenimis.

Visi programavimo darbai atlikti su Python programavimo kalba. RD kiekybiniai rodikliai buvo apskaičiuoti 8, 16, 32 s trukmės EEG fragmentams, kuriuose buvo pikai. Visais atvejais,

atlikta rodiklių statistinė analizė parodė skirtumus tarp dviejų pacientų grupių. Pacientų su smegenų pažeidimais kiekybinės analizės rodiklių aprašomosios charakteristikos (kvantilės, mediana, standartinis nuokrypis) buvo didesnės, palyginus su rodiklių aprašomosiomis charakteristikomis Rolando epilepsijos atveju. Kiekybinės analizės rodiklių skirtumai tarp dviejų pacientų grupių buvo pastebėti ir kanaluose, kuriuose nėra pikų.

16 s ilgio EEG fragmentų su pikais atveju, su atraminių vektorių klasifikatoriumi buvo pasiektas 72% klasifikavimo tikslumas, testavimui naudojant apmokyje nedalyvavusių pacientų duomenis. Neuroninio tinklo tikslumas svyravo 63-70% ribose. 8, 32 s EEG fragmentų ilgio atveju, klasifikavimo tikslumas buvo mažesnis dėl testinių duomenų specifikos. Atlikus tyrimą su skirtingu pikų kiekiu viename signale, nepavyko nustatyti priklausomybės tarp pikų kiekio signale ir klasifikavimo tikslumo.

Darbas yra tęstinis, todėl dalis skyrių medžiagos buvo panaudota iš praėjusiam semestre rašyto "Mokslo tiriamasis darbas (Mokslo tiriamojo darbo projektas)" darbo. Iš praeito semestro darbo buvo paimta sekančių skyrių medžiaga:

- 1. Rekurentinės diagramos.
- 2. Rekurentinių diagramų kiekybinė analizė.
- 4. Atraminių vektorių klasifikatorius.
- 6.1. Programavimo kalbos pasirinkimas.
- 6.3.2 50 Hz dažnio artefakto pašalinimas iš EEG signalo.

3, 6.2 skyrių medžiaga yra papildyta.

1. Rekurentinės diagramos

Norėdami suprasti mus supantį pasaulį, dažnai bandome atrasti įvairius procesuose slypinčius dėsningumus. Dauguma aplink mus vykstančių procesų yra netiesiniai ir chaotiški, juos veikia daugybė tarpusavyje susijusių faktorių. Viena iš fundamentalių tokių procesų savybių yra jų pasikartojamumas, t. y. procesas grįžta į savo pradinę ar artimą jai buseną. Šią savybę dar 1890 aprašė žymus prancūzų matematikas Anri Puankarė. Įrankis leidžiantis tirti proceso periodiškumą yra rekurentinė diagrama. Šis įrankis leidžia atskleisti sudėtingą proceso dinamiką paprasto dvimačio vaizdo pagalba. Analizuodami rekurentines diagramas galime atskleisti signalo pasikartojimo būdį, ar signalas yra generuotas chaotiško, o gal atsitiktinio proceso. Rekurentinė diagrama yra savotiškas signalo "piršto antspaudas", jos pagalba galime klasifikuoti įvairius signalus.

1.1. Rekurentinių diagramų apibrėžimas

Nagrinėjamas reiškinys apibrėžtas skaitmeniniu signalu su $N + 1$ reikšmiu

$$f_0, f_1, \dots, f_N.$$

Sugrupuokime turimą signalą po dvi reikšmes

$$y_0 = (f_0, f_1), y_1 = (f_1, f_2), \dots, y_{N-1} = (f_{N-1}, f_N).$$

Atlikus šią transformaciją, gavome signalų būsenų seką. Plačiau apie signalo būsenas kalbėsime sekančiame skyrelyje. Turėdami būsenų seką galime konstruoti rekurentinę diagramą. Tam tikslui palyginkime visas y_i reikšmes. Sakysime, kad jei signalo būsenos y_i ir y_j lygios ar beveik lygios tarpusavyje, tai diskretaus laiko indeksai i ir j žymės pasikartojančios būsenos laiką. Išivaizduokime koordinačių sistemą. Ant ašių atidėkime i ir j indeksus. Juodas taškas (i, j) žymės signalo būsenos pasikartojimą, t. y. būseną i panaši į būseną j , o baltas taškas reikš, kad (i, j) būsenos nėra tarpusavyje panašios. Toks juodai baltas taškinis paveikslėlis (angl. *bitmap*) vadinamas rekurentine diagrama.

Skaitytojui gali kilti klausimas, ką reiškia lygios ar beveik lygios būsenos. Būsenų vektorių porą $y_i y_j$ vadinsime panašia, jei atstumas tarp šių vektorių bus mažesnis nei tam tikras slenkstinis atstumas r . Matematiškai tai užrašoma sekančiu būdu:

$$\|y_i - y_j\| \leq r, i, j = 0, 1, \dots, M, r > 0, \quad (1.1)$$

kur M yra būsenų vektorių skaičius, r bet koks skaičius pasirinktas vartotojo. Taigi, jei atstumas tarp vektorių y_i ir y_j yra mažesnis už slenkstį r , tai taškas (i, j) bus juodos spalvos, priešingu atveju - baltos.

1.2. Signalo būsenų sąvoka

Praėjusiame skyrelyje pavartojome sąvoką signalo būseną. y_i vektorių konstavimas iš pirmo žvilgsnio gali pasirodyti perteklinis, juk galima gauti signalo rekurentinę diagramą tiesiogiai nau-

dojant signalo reišmes. Tačiau toks tiesmukas metodo taikymas nebūtų visiškai korektiškas. Dažnai nagrinėjamas signalas yra tik vienmatė sudėtingo proceso reprezentacija. Realaus pasaulio procesai priklauso nuo daugelio parametrų, todėl vienos dimensijos signalo atvaizdavimas rekurentine diagrama neparodytų tikrojo signalo pasikartojimų pobūdžio. Idealiu atveju, turėtumėme stebėti visus reiškinių įtakojančius veiksnius ir sudarydami būsenų vektorius iš šių reikšmių, konstruoti rekurentines diagramas. Tačiau, realybėje to padaryti neįmanoma. Takens 1981 metais pristatė teoremą, kurioje teigiama, jog dinaminės sistemos, kuriai įtakos turi daug kintamųjų, savybes galima atsatyti stebint tik vieną tos sistemos kintamąjį [33] panaudojant laiko vėlinimo metodą (angl. *method of time delays*). Iliustruokime šią teoremą pavyzdžiu. Nagrinėkime sekančią dinaminę sistemą:

$$X_{i+1} = Y_i + 1.0 - 1.4X_i^2 \quad (1.2)$$

$$Y_{i+1} = 0.3X_i. \quad (1.3)$$

Lygtys (1.2) ir (1.3) apibrėžia Henono sistemą, turinčią keistąjį atraktorių [33]. Atraktorius - aibė padėčių (taškų) kuriais juda dinaminė sistema praėjus pakankamai daug laiko. Tai yra, taškai, atsidūrę arti atraktoriaus ten ir išliks, net jei jų judėjimą šiek tiek sutrikdysime. Geometrine prasme atraktorius gali būti taškas, kreivė, daugdara arba taškų/kreivių aibė su fraktaline struktūra – keistasis atraktorius [1]. 1 pav. pavaizduotas Henono keistojo atraktoriaus atkūrimas. Tam tikslui buvo sugeneruotos 1200 reikšmės turinčios X_i ir Y_i eilutės, pradinėmis reikšmėmis pasirinkus nulius. Paėmus paskutines 200 šių eilučių reikšmių, X_i ir Y_i poros buvo pavaizduotos 1 pav. kairėje, 1 pav. viduryje buvo pavaizduotos X_{i+1} ir X_i poros, dešinėje - Y_{i+1} ir Y_i poros. Nors sistemą apibrėžia kintamieji X ir Y , tačiau kaip matyti iš 1 pav. vidurinėsios ir dešinės dalių, atraktoriaus formą galima atkūrti turint tik vieną šios sistemos kintamąjį.

Taigi, turėdami tik vieną diaugiadimensinės sistemos kintamąjį, galime atkurti šios sistemos ypatybes. Bendru atveju, signalo būsenų vektorius gaunamas sekančiu būdu:

$$y_i = (f_{i,i+d}, f_{i,i+2d}, \dots, f_{i,i+(D-1)d}), i = 0, 1, \dots, M, M = N - (D - 1)d. \quad (1.4)$$

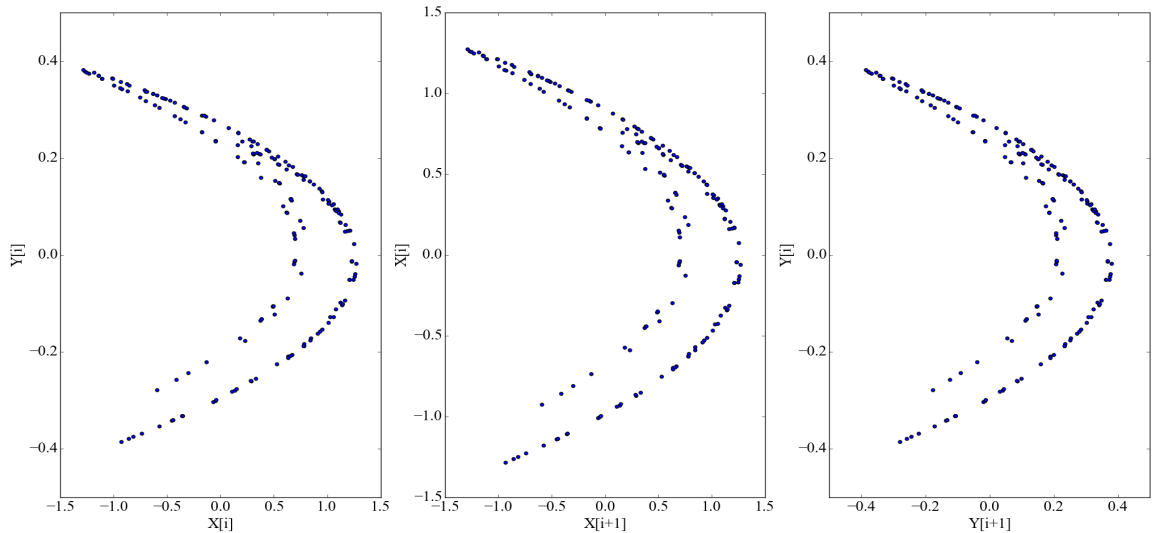
Šio vektoriaus dydis yra D . Signalo reikšmių pasirinkimas ne paeiliui, o su tam tikru vėlinimo parametru d yra susijęs su chaoso teorija. Mažai susiję tarpusavyje taškai leidžia tiksliausiai atkurti chaotiškos sistemos atraktorių. Apie parametrų D ir d kalbėsime 1.4 skyriuje.

1.3. Atstumo tarp dviejų būsenų skaičiavimas

Šiame skyriuje bus aptartas atstumo tarp dviejų būsenų skaičiavimas. Atstumas tarp dviejų vektorių y_i ir y_j žymimas $\|y_i - y_j\|$ ir apibrėžiamas funkcija, kuri vadinama metrika arba norma.

Dažniausiai naudojama norma yra Euklidinė metrika. Jos pagalba apibrėžtas atstumas tarp dviejų D - dimensinių vektorių atrodo sekančiu būdu:

$$\|y_i - y_j\| = \sqrt{(y_{i(1)} - y_{j(1)})^2 + (y_{i(2)} - y_{j(2)})^2 + \dots + (y_{i(D)} - y_{j(D)})^2}$$



1 pav. Henono sistemos keistasis atraktorius. Kairėje - atraktorius gautas atidedant y_i reišmes x_i atžvilgiu, viduryje - x_i x_{i+1} atžvilgiu, dešinėje - y_i y_{i+1} atžvilgiu. Gauto atraktoriaus forma identiška visais atvejais.

Dvimatėje ir trimatėje erdvėje ši formulė išvedama remiantis Pitagoro teorema. Literatūroje ši metrika vadinama L_2 . Kita galima metrika yra Manheteno metrika. Ji apskaičiuojama pasinaudojus šia formule:

$$\|y_i - y_j\| = |y_{i(1)} - y_{j(1)}| + |y_{i(2)} - y_{j(2)}| + \dots + |y_{i(D)} - y_{j(D)}|$$

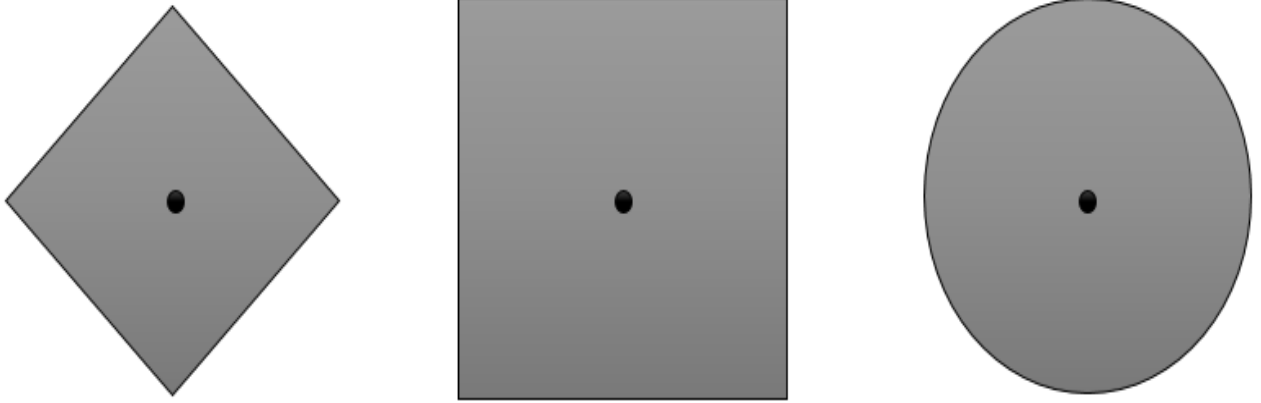
Ši metrika taip pat vadinama L_1 . Trečia, rekurentinių diagramų skaičiavime naudojama metrika, yra maksimumo metrika. Jos pavidalas yra sekantis:

$$\|y_i - y_j\| = \max(|y_{i(1)} - y_{j(1)}|, |y_{i(2)} - y_{j(2)}|, \dots, |y_{i(D)} - y_{j(D)}|)$$

Ji žymima L_∞ . Ši norma yra pigiausia procesoriaus darbo atžvilgiu. Tuo tarpu Euklidinė yra brangiausia. Nuo metrikos pasirinkimo priklauso juodų taškų kiekis rekurentinėje diagramoje. Esant vienodam slenkstiniam atstumui, daugiausiai juodų taškų turės diagrama gaunama su maksimumo norma, mažiausiai - su Manheteno. Euklidinė norma yra tašku kiekiu bus tarp maksimumo ir Manheteno. Tą iliustruoja 2 pav.. Jame pavaizduota visų metrių grafinė iliustracija dvimatėje erdvėje. Galime pastebėti, jog didžiasias aplinkos plotas yra gaunamas su maksimumo norma, mažiausias - su Manheteno. Juodų taškų kiekis visuose rekurentinėse diagramose skirsis neženkliai ir tai, savo ruožtu, neįtakos rekurentinių diagramų interpretacijos.

1.4. Parametrų pasirinkimas

Norėdami gauti kokybiškas rekurentines diagramas, turime tinkamai pasirinkti rekurentinės diagramos parametrus. Svarbiausi parametrai yra būsenų dimensija D , vėlinimas d ir slenkstinis atstumas r .



2 pav. Trys metrikos vaizduojančios to paties spindulio aplinką (pilkas plotas) aplink juodą tašką dvimatėje erdvėje. Kairėje - Manheteno metrika, viduryje - maksimumo metrika, dešinėje - Euklidinė metrika.

1.4.1. Būsenų dimensijos pasirinkimas

Būsenų dimensija D parodo, kokio sudėtingumo yra nagrinėjamas procesas. Signalo dimensiją apytiksliai galime nustatyti skaičiuodami koreliacinę dimensiją.

Koreliacinė dimensija tampriai susijusi koreliacinio integralo sąvoka. Koreliacinis intergalas $C(r)$ tai juodų taškų santykis su bendru tašku kiekiu rekurentinėje diagramoje. Į šį santykį neįeina trivialūs juodi taškai, kurie yra rekurentinės diagramos pagrindinėje įstrižainėje.

Iš koreliacinio integralo $C(r)$ ir r priklausomybės galima gauti informacijos apie koreliacinę dimensiją D_c [18]. Signalo koreliacinis integralas yra proporcingas koreliacinei dimensijai D_c (su kažkokia proporcingumo konstanta β):

$$C(r) = \beta r^{D_c},$$

arba

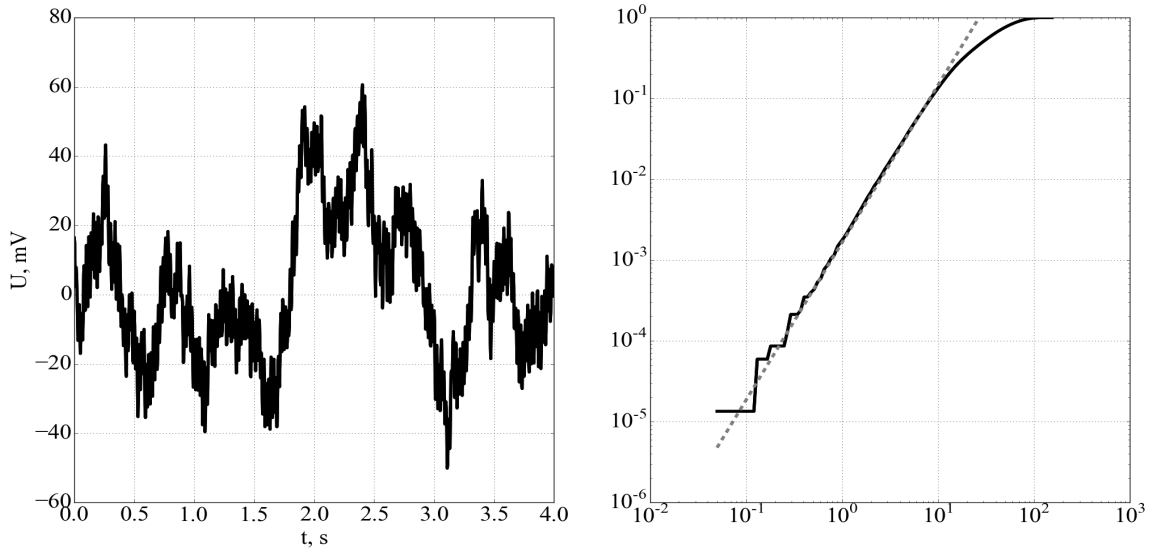
$$\log_{10} C(r) = D_c \log_{10} r + \log_{10} \beta. \quad (1.5)$$

Atkreipkime dėmesį, jog logaritmas pagrindu 10 1.5 lygtyje yra naudojamas tik patogumo sumetimais. Pažymėję $x = \log_{10} r$, $y = \log_{10} C(r)$, $m = \log_{10} \beta$ gausime lygtį

$$y = D_c x + m. \quad (1.6)$$

Šios lygties krypties koeficientas ir bus ieškomas dydis D_c .

Praktiškai tai reiškia, jog turime nubraižyti $\log_{10} C(r)$ ir $\log_{10} r$ priklausomybę ir įvertinti nuolydžio koeficientą. 3 pav. parodytas D_c radimo algoritmas. Su $D = 2$ ir $d = 1$ apskaičiuota žmogaus elektrokardiogramos koreliacinė dimensija lygi 1.42. Skaitytojas gali pastebėti, jog D_c radime dalyvauja mūsų ieškomas dydis D . Iškyla vištos ir kiaušinio problema. Literatūroje siūloma šią problemą išspręsti iteraciniu būdu, t. y. didinti D reikšmę tol, kol D_c reikšmė įsisotina.



3 pav. EEG signalas (kairėje) ir šio signalo koreliacinio integralo $C(r)$ ir slenksčio parametro r priklausomybė (dešinėje, log-log skalė, ordinačių ašyje - koreliacinis integralas $C(r)$, abscisių - slenksčio parametras - r). Įvertinta koreliacinė dimensija $D_c = 1.95$, kai būsenų dimensija $D = 2$, vėlinimas $d = 1$.

Kaip matome, D_c nustatymas yra darbu imlus procesas. Be to, rezultatus gali iškreipti įvairaus pobūdžio triukšmai, kurių yra apstu medicininės kilmės signaluose. Todėl [33] darbe siūloma fiziologinės kilmės duomenims būsenų dimensiją nustatyti lygia 10.

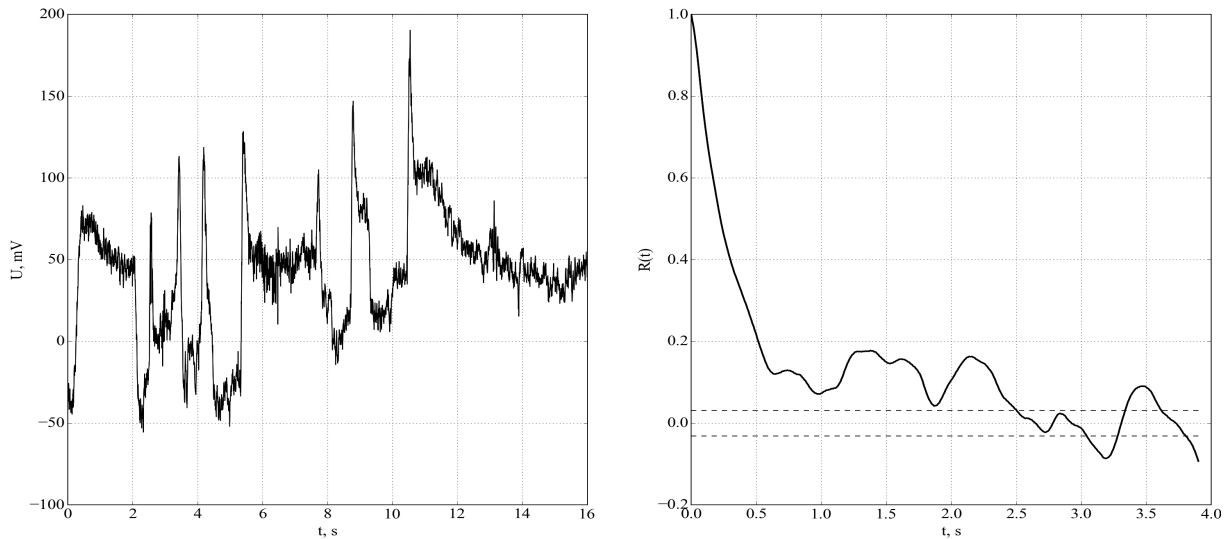
1.4.2. Vėlinimo parametro pasirinkimas

Vėlinimo parametras d nusako, koku atstumu vienas nuo kito yra 2 gretimi taškai būsenų vektoriuje. Parametras $d = 1$ reiškia, jog į būsenos vektorių pateks D paeiliui einančių signalo stebėjimų, $d = 2$ - kas antras ir t. t. Literatūroje rekomenduojama nustatyti d reikšmę tokią, sulyg kurią autokoreliacijos funkcija (1.7 formulė), pirmą kartą tampa artima nuliui, t. y. jos modulis neviršija $2/\sqrt{N}$. 5 pav. parodyta vieno iš EEG kanalų autokoreliacijos funkcija. Akreipkime dėmesį, jog pav. pavaizduota ne $r(d)$ priklausomybė, tačiau $R(t)$. Iš tiesų $r(d) = R(t)$ tik vietoje d naudojama transformacija $t = d\Delta t$, kur Δt yra signalo f diskretizavimo žingsnis (EEG signalo diskretizavimo dažnis 256 Hz, o žingsnis $\Delta t = \frac{1}{256}$ (Plačiau žr. 3.2 skyrelį). Iš pav. matyti, kad EEG signalo autokoreliacijos funkcija pirmą kartą priartėja prie nulio esant $t = 2.5s$, kas atitiktų $d = 640$. [33] darbe nurodoma, jog vėlinimo parametras d nėra esminis, todėl šiame darbe bus naudojama jo reikšmė lygi 1.

$$r(d) = \frac{\sum_{j=0}^{N-d} (f_j - \bar{f})(f_{d+j} - \bar{F})}{\sqrt{\sum_{j=0}^{N-d} (f_j - \bar{f})^2 \sum_{j=0}^{N-d} (f_{d+j} - \bar{F})^2}}, d = 0, 1, \dots [N/2] \quad (1.7)$$

čia

$$\bar{f} = \frac{1}{N-d+1} \sum_{i=0}^{N-d} f_i, \bar{F} = \frac{1}{N-d+1} \sum_{i=0}^{N-d} f_{d+i}. \quad (1.8)$$

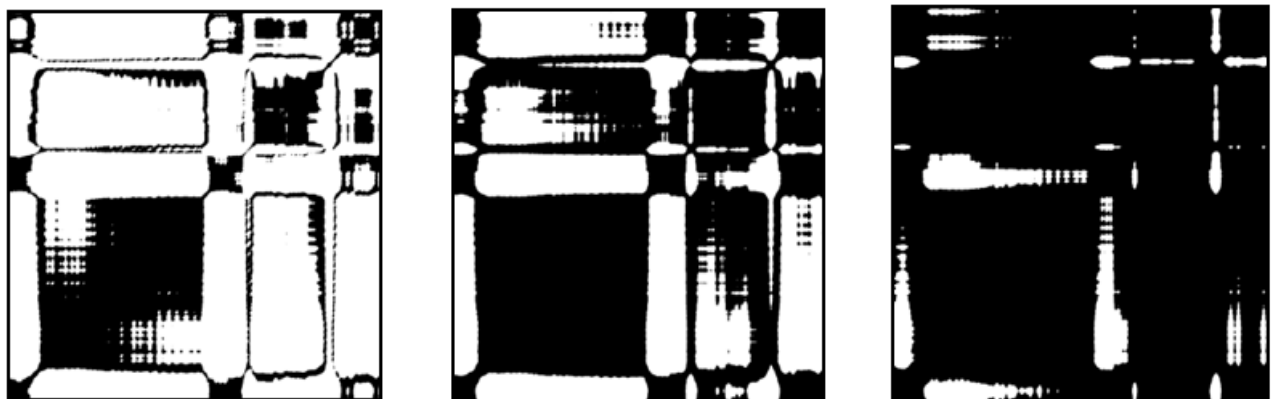


4 pav. Vieno iš EEG kanalų fragmentas ir jo autokoreliacijos funkcija. Autokoreliacijos funkcija pateikta sekundėmis. Signalas priartėja prie nulio po 2.5 s, t.y. $d = \frac{t}{\Delta t} = 640$.

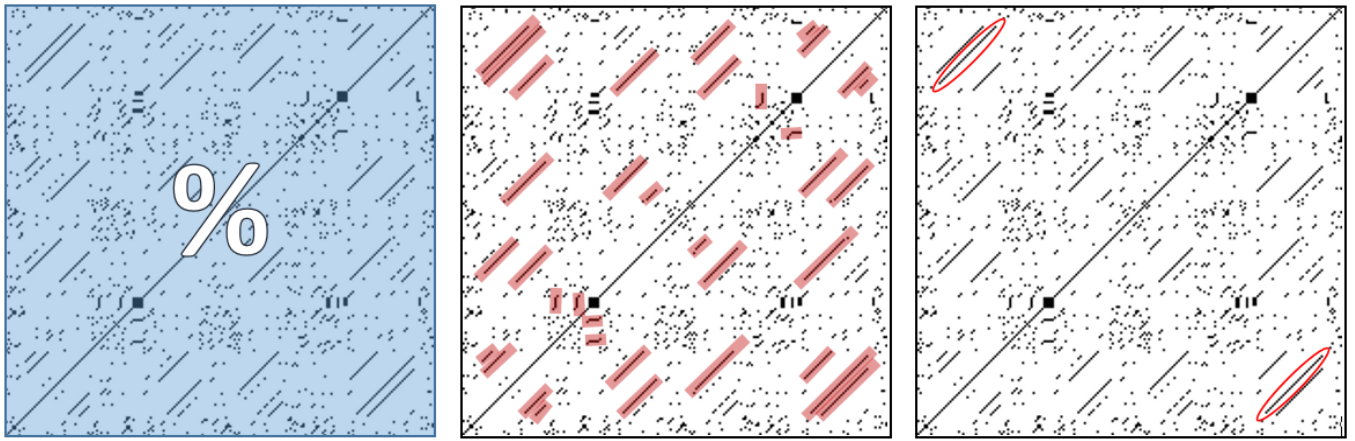
1.4.3. Slenkstinio atstumo pasirinkimas

Nuo slenkstinio atstumo r priklauso juodų taškų kiekis rekurentinėje diagramoje. Mažindami r gausime vis mažiau juodų taškų. Kai $r \rightarrow 0$ juodi taškai išliks tik pagrindinėje įstrižainėje. Teoriškai, idealiai periodiniam signalui, net su $r = 0$ turėtų būti matomos įstrižos linijos nutolusios nuo pagrindinės įstrižainės per periodo ilgį, tačiau dėl kompiuterinių skaičiavimų paklaidų to dažnai neįvyksta. Didindami r gausime vis daugiau juodų taškų. Galiausiai r reikšmei viršyjus didžiausią atstumą tarp dviejų būsenų vektorių, gausime visiškai juodą rekurentinę diagramą.

Atkreipkime dėmesį, jog absoliuti r reikšmė matuojama tokiais pačiais matavimo vienetais, kaip ir nagrinėjamas signalas. Tačiau praktikoje dažniausiai naudojama santykinė r reikšmė išreikšta kaip procentas nuo maksimalaus atstumo tarp dviejų būsenų vektorių [33]. Rekomenduojama slenkstinį atstumą pasirinkti tokį, jog juodų taškų kiekis rekurentinėje diagramoje neviršytų 2% [33].



5 pav. Vieno iš EEG kanalų rekurentinės diagramos su skirtingomis slenkstinio atstumo r reikšmėmis. Kairėje - $r = 67.44$ (30% juodų taškų), centre - $r = 152.99$ (60% juodų taškų, dešinėje - $r = 281.67$ (90% juodų taškų). $D = 10$, $d = 1$.



6 pav. Rekurentinių diagramų kiekybinė analizė remiasi juodų tašku tankiu, įstrižomis bei vertikaliomis linijomis RD. Kadangi RD yra simetriška pagrindinės įstrižainės atžvilgiu, tai užtenka nagrinėti RD virš (žemiau) pagrindinės įstrižainės. Kairėje - juodų taškų tankis RR . Centre pavaizduotos įvairaus ilgio įstrižos linijos. Jos paryškintos pilka spalva. Rodikliai besiremiantys šiomis įstrižomis linijomis yra L , DET , $ENTR$. Dešinėje apbraukta ilgiausia įstriža linija (L_{max}). Šiame paveiksle neparodyti rodikliai besiremiantys vertikaliomis linijomis, tačiau jų apskaičiavimo principas yra identiškas. Rekurentinės diagramos gautos iš dirbtinio signalo.

2. Rekurentinių diagramų kiekybinė analizė

2.1. Rodikliai

Vizuali rekurentinių diagramų analizė gali atsakyti į klausimus apie signalo stacionarumą, periodiškumą ir atsitiktinumą, tačiau jos pagalba negalime tiksliai klasifikuoti signalus pagal jų sudėtingumą. Šiuo tikslu, darbuose [16], [34] buvo pasiūlyti rodikliai, leidžiantys vertinti signalų sudėtingumą.

Rodikliai remiasi juodų taškų tankiu bei įstrižų ir vertikalų linijų struktūromis RD. Juodi taškai rekurentinėje diagramoje parodo būsenų pasikartojimą. Įstrižos linijos vaizduoja signalo kvazi-periodiškumo ruožus signale. Kuo ilgesnė įstriža linija, tuo ilgiau signalas buvo periodinis. Idealiai periodinio signalo atveju, rekurentinė diagrama susidarys vien iš įstrižų linijų, lygiagrečių pagrindinei įstrižainei. Pagrindinė įstrižainė neturi jokios informacijos apie signalą, nes ji yra sudaryta iš trivialių taškų, t. y. taškų, kurie rodo būsenų panašumą pačių į save. Vertikalios bei horizontalios linijos signalizuoja apie nusistovėjusių būsenų būvimą signale, t. y. tam tikrais laiko tarpais signalas beveik arba visiškai nekinta. 2 paveiksle parodytas schematinis kiekybinis rodiklių vaizdas. Įstrižos bei vertikalios linijos paryškintos raudonai.

Toliau bus pristatyti šie rodikliai.

2.1.1. Rodikliai besiremiantys juodų taškų tankiu

Pats paprasčiausias kiekybinės analizės rodiklis yra pasikartojimo norma (angl. *recurrence rate*). Jo išraiška

$$\%RR = 100 \frac{\#juodi_taskai_virsutiniame_trikampyje}{M(M-1)/2}, \quad (2.1)$$

kur M - RD dydis.

Šis rodiklis matuoja juodų taškų tankį rekurentinėje diagramoje. Kadangi RD yra simetriška pagrindinės įstrižainės atžvilgiu, todėl skaičiavimams yra naudojama informacija esanti virš (žemiau) pagrindinės įstrižainės.

2.1.2. Rodikliai besiremiantys įstrižomis linijomis

Sekantys rodikliai remiasi įstrižomis rekurentinės diagramos linijomis. Įstrižos linijos parodo signalo kvazi-periodiškumo epizodus.

Stochastinių arba chaotiškų procesų rekurentinėse diagramose įstrižų linijų skaičius būna mažesnis negu deterministinių procesų, t. y. šie procesai turi mažiau periodiškumo ar kvazi-periodiškumo ruožų. Santykis taškų formuojančių įstrižas linijas (minimalaus ilgio l_{min}) ir visų rekurentinės diagramos taškų vadinamas determinizmu (angl. *determinism*). Jo išraiška pavaizduota (2.2) lygtyje. Pasirinkus $l_{min} = 1$, determinizmas bus lygus 1.

$$\%DET = 100 \frac{\#juodi_taskai_istrizose_linijose}{\#juodu_tasku_kiekis} \quad (2.2)$$

Kitas rodiklis yra vidutinis įstrižos linijos ilgis (angl. *average diagonal line length*). Įstrižos linijos ilgis l reiškia, jog l kartų iš eilės signalo būsenos skirtingose atkarpose buvo panašios. Jo išraiška pavaizduota (2.3) lygtyje.

$$L = vidutinis_istrizos_linijos_ilgis \geq l_{min}, \quad (2.3)$$

kur l_{min} - minimalus įstrižos linijos ilgis.

L_{max} parodo kokia buvo ilgiausia įstriža linija aptikta rekurentinėje diagramoje.

$$L_{max} = ilgiausia_istriza_linija_rekurentineje_diagramoje \quad (2.4)$$

Entropijos (angl. *entropy*) rodiklis matuoja rekurentinės diagramos sudėtingumą atsižvelgiant į įstrižas linijas. Triukšmo entropijos rodiklis turėtų būti mažas, t.y. signalas nėra sudėtingas. Rodiklio išraiška pavaizduota 2.5 lygtyje.

$$ENTR = - \sum_{l=l_{min}}^{L_{max}} p(l) \ln p(l), \quad (2.5)$$

kur $p(l) = \frac{l_ilgio_istrizu_liniju_kiekis}{N_l}$ - tikimybė rasti l ilgio įstrižą liniją rekurentinėje diagramoje, N_l - įstrižų linijų, kur $l \geq l_{min}$, suma.

2.1.3. Rodikliai besiremiantys vertikaliomis linijomis

Kiti rodikliai naudojami kiekybinėje analizėje remiasi rekurentinės diagramos vertikaliomis linijomis. Vertikalios linijos rekurentinėje diagramoje nurodo, jog signale egzistuoja nusistovėjusios būsenos, t. y. signalo reikšmės nesikeičia bėgant laikui.

Pirmasis rodiklis besiremiantis vertikaliomis linijomis yra nusistovėjimas. Nusistovėjimas (angl. *laminarity*) - tai santykis juodų taškų formuojančių vertikalias linijas su visais juodais taškais. Parodo kiek signalas turi nusistovėjusių būsenų.

$$\%LAM = 100 \frac{\#juodi_taskai_vertikaliose_linijose}{\#juodu_tasku_kiekis} \quad (2.6)$$

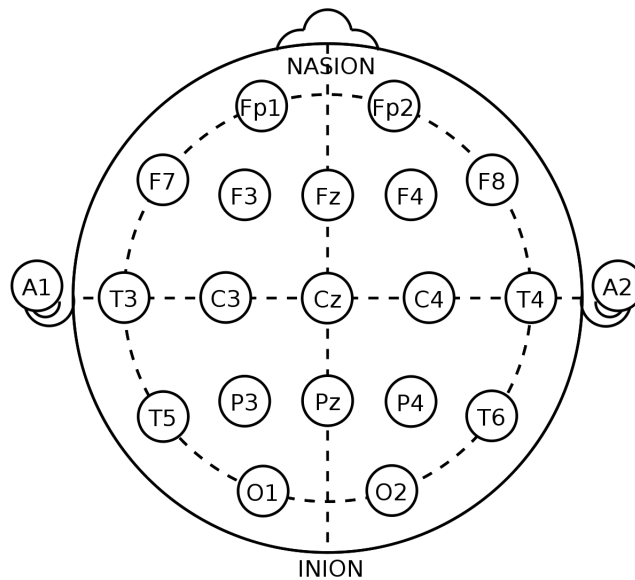
Kitas rodiklis yra nusistovėjimo laikas. Nusistovėjimo laikas (angl. *trapping time*) tai vidutinis vertikalių linijų ilgis. Jis parodo, kiek vidutiniškai procesas būna nusistovėjusioje būsenoje.

$$TT = vidutinis_vertikalios_linijos_ilgis \geq v_{min}, \quad (2.7)$$

kur v_{min} - minimalus vertikalių linijų ilgis.

V_{max} parodo kokia buvo ilgiausia įstriža linija aptikta rekurentinėje diagramoje.

$$V_{max} = ilgiausia_vertikali_linija_rekurentineje_diagramoje \quad (2.8)$$



7 pav. Standartinės 10-20 sistemos elektrodų išdėstymo schema. Šaltinis: [2]

3. Elektroencefalogramos

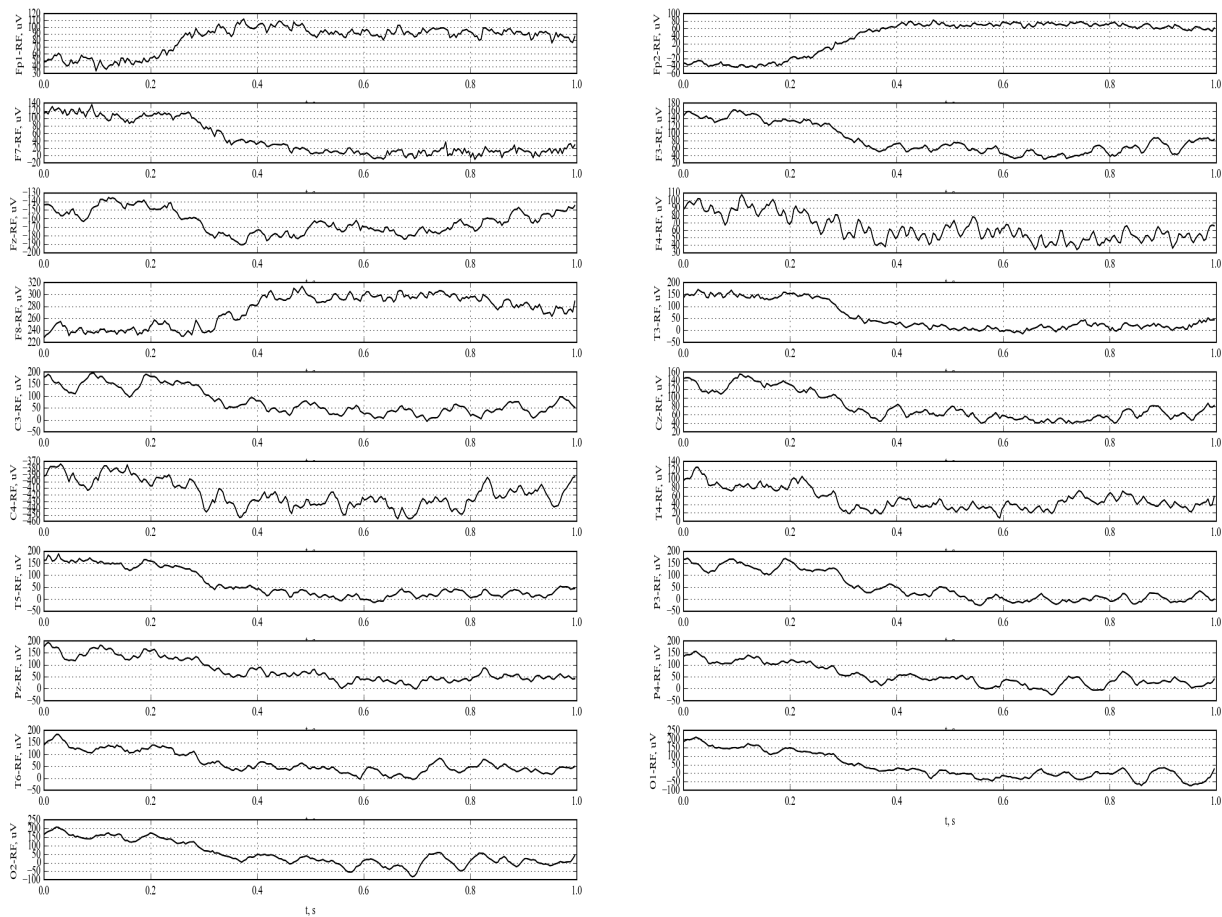
Šioje darbo dalyje bus aptarti EEG signalai. Bus aptartas EEG signalų gavimas, standartinė 10-20 sistema, taip pat EEG taikymas įvairių ligų diagnostikai bei EEG atsirandantys artefaktai.

3.1. EEG signalų gavimas

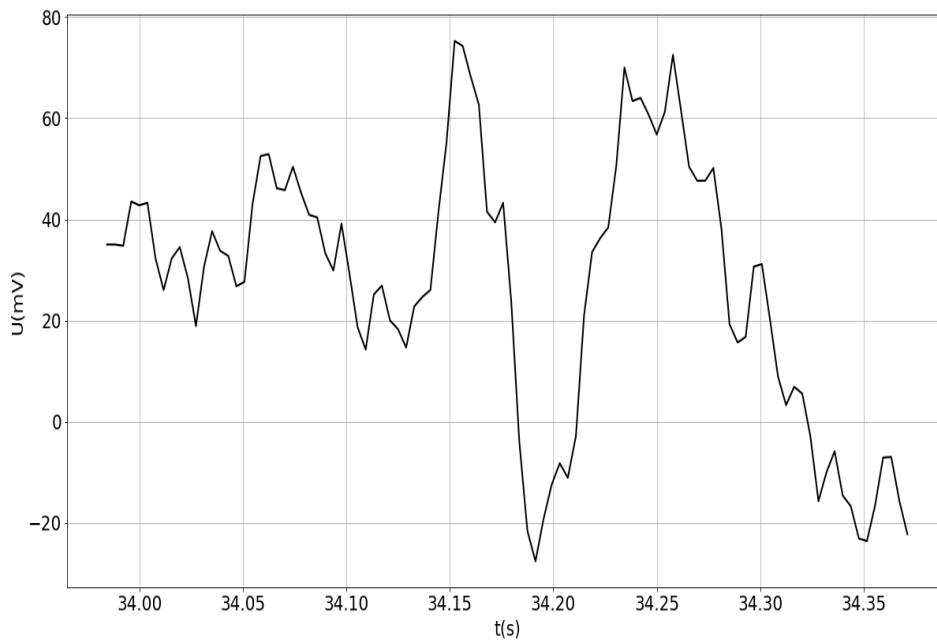
Smegenų veiklai yra būdingas elektrinis aktyvumas. Šio elektrinio aktyvumo priežastis - fiziologiniai procesai vykstantys sąveikaujant smegenų ląstelėms. Elektrinis smegenų aktyvumas išmatuojamas prie galvos pritvirtintais elektrodais. Gauti signalai yra mikrovoltų eilės. Tipinė EEG signalų amplitudė yra nuo $10 \mu\text{V}$ iki $100 \mu\text{V}$ ir dažnis nuo 1 Hz iki 100 Hz [29].

3.2. 10-20 sistema

10-20 sistema tai tarptautiniu lygiu pripažintas metodas elektrodų išdėstymui. Sistema paremta elektrodų tarpusavio išdėstymu ant galvos. Iš čia kyla ir sistemos pavadinimas. Atstumai tarp elektrodų būna 10 arba 20 procentų atstumo nuo kaktos ir pakaušio (galvos ilgis) arba atstumo tarp smilkinių (galvos plotis), todėl elektrodai kiekvienam pacientui būna apytiksliai virš tos pačios smegenų srities. Tokiu būdu užtikrinamas klinikinių testų atkartojamumas ir rezultatų palyginimas. 3.2 pav. pavaizduota 10-20 sistemos schema. 8 pav. pavaizduoti signalai gauti pasinaudojus 10-20 sistema. Kiekvienas elektrodas generuoja atskirą signalą. Analoginis signalas paverčiamas skaitmeniniu naudojantis konverteriais. Prasmingi EEG dažniai yra iki 100 Hz, todėl minimalus 200 stebėjimų per sekundę diskretizacijos dažnis (siekiant patenkinti Naikvisto teoremą) dažnai yra pakankamas [27].



8 pav. 19 EEG kanalų 1 sekundės fragmentas. Signalo diskretizacijos dažnis - 256 Hz.



9 pav. Paciento sergančio Rolando epilepsija T3 kanalo fragmentas. Pikas matomas ties 34.2 sekunde.

3.3. EEG taikymas diagnostikoje

EEG plačiai taikomas tokių ligų kaip epilepsija, cerebrinis paralyžius diagnostikai. Siekiant diagnozuoti šias ligas, yra ieškoma epileptoforminių pikų, kurie šiame darbe bus vadinami pikais. Piko pavyzdys pavaizduotas 9 pav. Pikas gali turėti tiek teigiamą, tiek neigiamą amplitudę. Piko trukmė neturėtų būti mažesnė negu 40 ms ir ne didesnė negu 200 ms [19]. Pikų paieška atliekama tiek rankiniu būdu, tiek automatizuotų algoritmų pagalba. Automatizuotų pikų paieškos algoritmų apžvalga pateikiama [12] darbe. Šiame darbe visi pikai buvo rasti algoritmo, kuris aprašytas [19] darbe. Šis algoritmas yra paremtas morfologiniais filtrais, kurių pagalba išfiltruojama foninė smegenų veikla, taip išryškinant pikų vietas.

EEG matavimai atliekami įvairiose paciento būsenose. Jie gali būti atlikti tiek budrumo būsenoje, tiek miego metu, taip pat esant užmerktoms arba atmerktoms akims.

3.4. EEG artefaktai

Nors EEG tikslas yra išmatuoti smegenų veiklą, dažnai kartu su smegenų veikla yra užfiksuojama ir kita veikla. Ši veikla yra vadinama EEG artefaktais. Artefaktai gali būti fiziologiniai arba išoriniai. Fiziologinės kilmės artefaktai yra generuojami paciento kūno. Tai yra širdies veiklos artefaktai, raumenų artefaktai bei akių judesių artefaktai. Išoriniai artefaktai susiję su įvairių išorinių prietaisų sukeltais artefaktais bei pačios įrašymo aparatūros netobulumais. Dažniausias išorinio artefakto tipas yra 50/60 Hz dažnio elektros tinklo artefaktas. Taip pat artefaktus gali sukelti telefonai ir kiti prietaisai skleidžiantys elektromagnetines bangas.

4. Atraminių vektorių klasifikatorius

Šiame skyriuje bus aptartas atraminių vektorių klasifikatorius. Skyrius remiasi [30] knygoje pateikta medžiaga.

Atraminių vektorių klasifikatorius (angl. *support vector machine, SVM*) yra galingas mašininio mokymosi metodas. Jis yra naudojamas klasifikavimui ir regresinei analizei. Metodo esmė yra surasti hiperplokštumą, kuri geriausiai atskiria analizuojamas klases. Šiuo atveju geriausiai atskirianti tiriamas klases hiperplokštuma yra ta, kurios atstumas iki artimiausių klasių atstovų, kurie dar yra vadinami atraminiais vektoriais, yra maksimalus. Hiperplokštumos lygtis, kuri atskiria nagrinėjamas klases atrodo sekančiu būdu:

$$\vec{w} \cdot \vec{x} + b = 0, \quad (4.1)$$

kur \vec{w} ir b - modelio parametrai.

Pažymėję klases $+1$ ir -1 , naujo stebėjimo \vec{z} klasę y galėsime nustatyti atsižvelgus į jo padėtį hiperplokštumos atžvilgiu, sekančiu būdu:

$$y = \begin{cases} 1, & \text{jei } \vec{w} \cdot \vec{z} + b > 0; \\ -1, & \text{jei } \vec{w} \cdot \vec{z} + b < 0. \end{cases} \quad (4.2)$$

Modelio parametras \vec{w} surandamas išsprendus šią optimizavimo problemą:

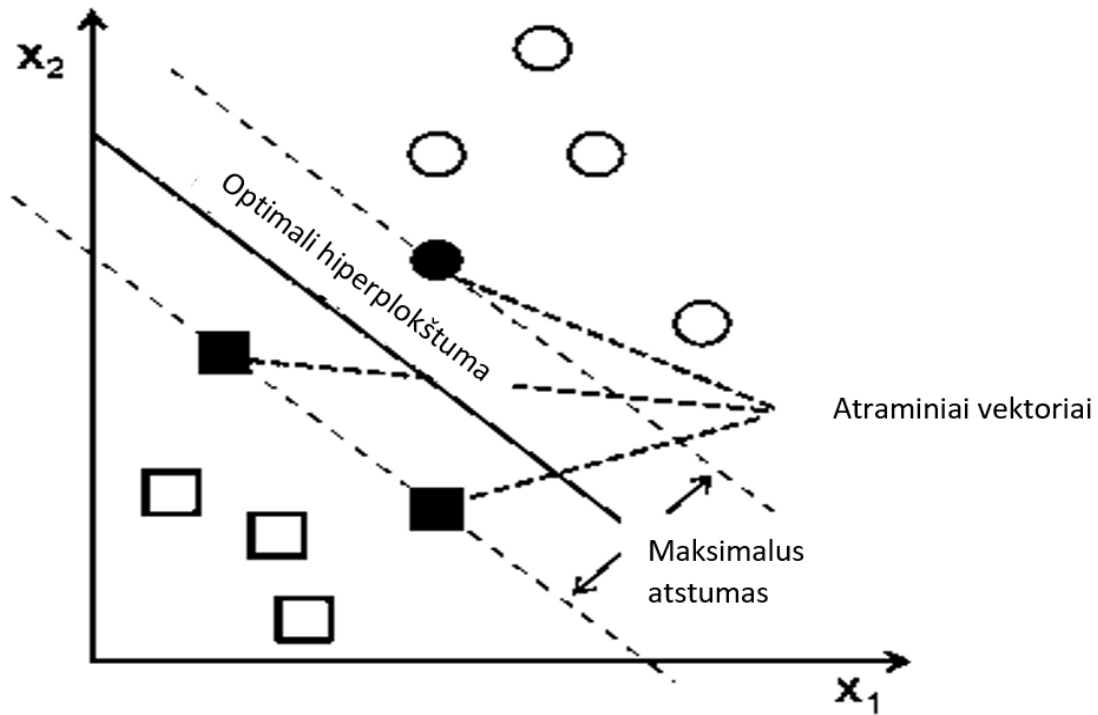
$$\begin{aligned} \min_w \frac{\|\vec{w}\|^2}{2} \\ y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, i = 1, \dots, N. \end{aligned} \quad (4.3)$$

Atraminių vektorių klasifikatorius pagal apibrėžimą stengiasi surasti tokią hiperplokštumą, kuri idealiai atskiria nagrinėjamas klases. Akivaizdu, jog duomenys ne visada yra lengvai atskiriami. Kai kuriais atvejais duomenų taškai gali patekti į "tarpą" tarp klasių. Tuo tikslu yra naudojama "minkštos" ribos (angl. *soft margin*) technika, kuri leidžia nagrinėjamoms klasėms persidengti, bet už tai atitinkamai baudžia. Toliau pateikiama modifikuota 4.3 tikslo funkcija, leidžianti kai kuriems klasės taškams persidengti:

$$\begin{aligned} \min_w \frac{\|\vec{w}\|^2}{2} + C \sum_i \xi_i \\ y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi_i, i = 1, \dots, N. \end{aligned} \quad (4.4)$$

Čia ξ_i - "laisvumo" kintamasis (angl. *slack variable*). Kintamasis leidžia duomenų taškams išėiti iš nustatytų ribų, tačiau, juos atitinkamai baudžia. Algoritmas stengiasi išlaikyti "laisvumo" kintamojo reikšmę artimą 0, tuo pačiu stengiasi surasti hiperplokštumą su maksimaliu atstumu nuo klasių. Parametras C padeda surasti optimalų dydį \vec{w} atsižvelgiant į nekorektiško klasifikavimo atvejus. Galimi keli C parametro pasirinkimai:

- Maža C reikšmė leidžia beveik neatsižvelgti į neatitikimus, kas savo ruožtu įgalina nustatyti platesnę ribą tarp klasių;



10 pav. Atraminų vektorių klasifikatorius atskiria nagrinėjamas klases (pažymėti apskritimais ir kvadratais) surasdama hiperplokštumą (juoda linija), kuri yra maksimaliai nutolusi nuo atraminų vektorių. Šaltinis: [5].

- Didelė C reikšmė neleidžia ignoruoti neatitikimų, todėl nustatoma siauresnė riba tarp klasių;
- Kai $C = \inf$ (labai didelis skaičius) neatitikimai neleidžiami.

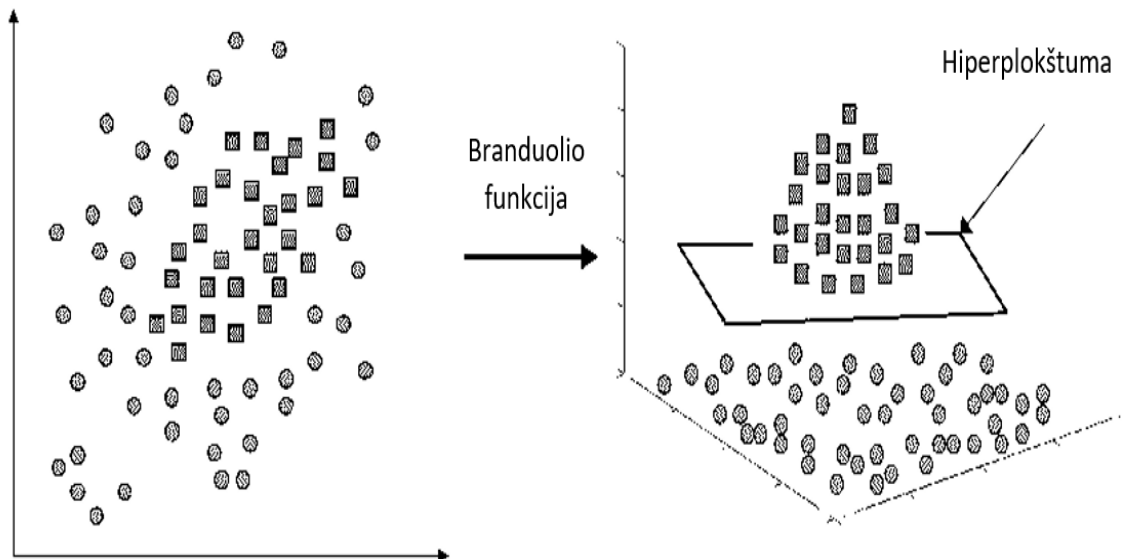
Gan dažnai duomenys pasižymi netiesiškumu ir juos neįmanoma atskirti hiperplokštuma. Norint atskirti tokius duomenis, reikia naudoti branduolio funkciją (angl. *kernel function*), kuri transformuoja pradinis duomenis į didesnio matmenų erdvę, taip atskiriant duomenis (žr. 11 pav.). Toliau bus pavaizduotos 4 pagrindinės branduolio funkcijos [13]:

- tiesinė: $K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j$;
- polinominė: $K(\vec{x}_i, \vec{x}_j) = (\gamma \vec{x}_i \cdot \vec{x}_j + r)^d, \gamma > 0$;
- radialinių bazių funkcija (RBF): $K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2), \gamma > 0$;
- sigmoidinė: $K(\vec{x}_i, \vec{x}_j) = \tanh(\gamma \vec{x}_i \cdot \vec{x}_j + r)$.

Čia γ, r, d - branduolio funkcijų parametrai. Šiame darbe bus naudojama RBF branduolio funkcija.

4.1. Parametrų pasirinkimas

Nuo tinkamų parametrų pasirinkimo priklauso SVM klasifikavimo kokybė. Radialinių bazių funkcija turi 2 parametrus: C ir γ . Iš anksto nėra žinoma, kokie parametrai konkrečiai problemai yra geriausi. Todėl yra būtina atlikti tam tikrą parametrų paiešką. Tikslas yra surasti tokį derinį



11 pav. Kai duomenys pasižymi netiesiškumu, atraminių vektorių klasifikatorius naudoja branduolio funkciją atvaizduojant duomenis i didesnių matmenų erdvę. Šaltinis: [4].

$(C; \gamma)$, su kuriais testinių duomenų klasifikavimo tikslumas būtų geriausias. [13] straipsnio autoriai parametru paieškai rekomenduoja naudoti tinklo paiešką (angl. *grid-search*). Šio metodo esmė yra perrinkti visas $(C; \gamma)$ kombinacijas iš apibrėžtų sekų surandant porą pasižyminčia geriausiu klasifikavimo tikslumu. Klasifikavimo tikslumui nustatyti naudojamas v -sluoksnių kryžminis patikrinimas. Duomenys yra padalinami į v lygių dalių, kur viena dalis yra naudojama testavimui, likusios $v - 1$ - apmokymui. Procesas kartojamas v kartų, kad klasifikatoriaus tikslumas būtų patikrintas su kiekviena duomenų dalimi. [13] straipsnio autoriai rekomenduoja naudoti eksponentiškai augančias sekas (pvz. $C = 2^{-5}, 2^{-3}, \dots, 2^{15}; \gamma = 2^{-15}, 2^{-13}, \dots, 2^3$). Akivaizdus šio metodo trūkumas yra jo našumas, nes tenka perrinkti visas parametru kombinacijas. Nepaisant šio trūkumo, šis metodas dėl savo paprastumo yra plačiai naudojamas.

[8] straipsnio autoriai siūlo kitoki parametru optimizavimo algoritmą. Parametru paieškai jie siūlo naudoti atsitiktinę paiešką (angl. *random search*). Autoriai parodo, jog šis metodas gali būti toks pat efektyvus kaip ir tinklo paieška, bet reikalauti mažiau skaičiavimo resursų.

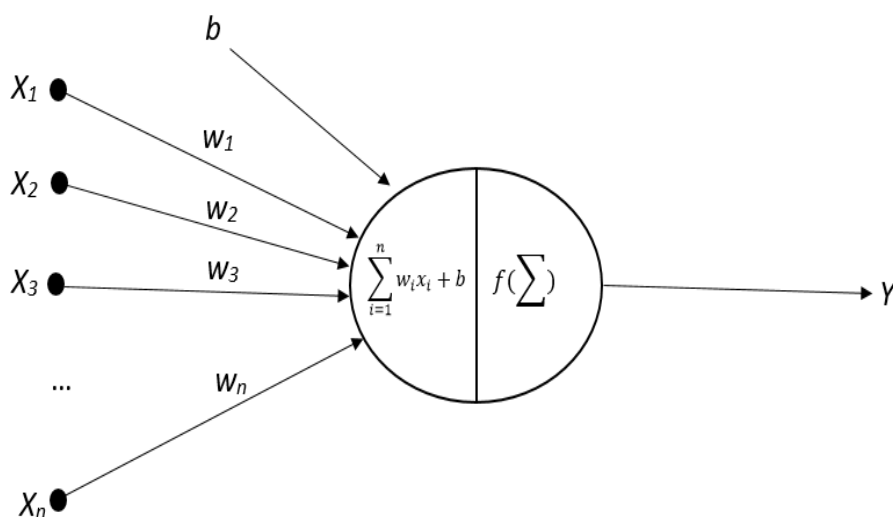
Šiame darbe gerų parametru paieškai bus naudojama tinklo paieškos metodas.

5. Dirbtiniai neuroniniai tinklai

Šiame skyriuje bus pristatyti dirbtiniai neuroniniai tinklai. Šio skyriaus informacija remiasi [32], [22] knygoje pateikta medžiaga.

Dirbtiniai neuroniniai tinklai yra bandymas sumodeliuoti biologines neuronų sistemas.

Žmogaus smegenys susideda apytiksliai iš 100 milijardų ląstelių. Šios ląstelės vadinamos neuronais. Neuronas susideda iš ląstelės kūno (somos), įėjimo ataugų dendritų ir ilgos išėjimo ataugos aksono. Vienų neuronų išėjimo ataugos yra sujungtos su kitų neuronų įėjimo ataugomis. Šios jungtys yra vadinamos sinapsėmis. Šiomis jungtimis yra perduodami elektrocheminiai signalai. Gavęs tam tikro stiprumo signalą, neuronas sužadinas ir perduoda signalą toliau. Ar bus perduotas signalas toliau priklauso nuo neurono sužadavimo slenksčio. Neurono sužadavimo tikimybė ir priimamo signalo stiprumas priklauso nuo sinapsių perdavimo efektyvumo. Mokslininkai mano, jog mokymosi procesas vyksta keičiantis sinapsių efektyvumui [32]. Dirbtinis neuroninis tinklas (DNT) - tai sistema skirta klasifikavimo ir regresijos uždaviniams spręsti. Dirbtiniai neuroniniai tinklai susideda iš daugelio tarpusavyje sujungtų paprastų skaičiavimo elementų - perceptronų. Šiuo metu sudėtingiausi neuroniniai tinklai gali būti sudaryti iš keliolikos milijardų perceptronų.



12 pav. Perceptrono modelis.

5.1. Perceptronas

Paprasčiausias neuroninio tinklo modelis yra sudarytas iš vieno skaičiavimo elemento - perceptrono. Šis tinklas dar vadinamas viensluoksniu perceptronu. Perceptronas susideda iš įėjimų, skaičiavimo elemento ir išėjimo. Modelio įėjimų kiekis atitinka duomenų atributų kiekį. Kiekvienam įėjimui yra priskirtas svoris w . Svorių pagalba įėjimai yra stiprinami arba slopinami. Modelis taip pat turi poslinkio (angl. *bias*) koeficientą, kuris įgalina modelį lanksčiau prisiderinti prie nagrinėjamų duomenų. Svoriai ir poslinkio koeficientas paprastai yra realieji skaičiai. Perceptrono struktūra pavaizduota 12 pav. Įėjimų vertės x_1, x_2, \dots, x_n yra padauginamos iš atitinkamų svorių w_1, w_2, \dots, w_n ir yra susumuojamos. Prie gautos sumos yra pridamas poslinkio koeficientas b .

Gauta suma yra perduodamos funkcijai $f()$. Funkcija $f()$ yra vadinama aktyvacijos funkcija. Nuo jos reikšmės priklauso perceptrono išėjimas y . Formaliai modelis aprašomas sekančiu būdu:

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right), \quad (5.1)$$

kur n - įėjimų skaičius.

5.2. Aktyvacijos funkcijos

Aktyvacijos arba perdavimo funkcija nusako ryšį tarp perceptrono įvesties ir išvesties. Dėl pasirinktos aktyvacijos funkcijos savybių, perceptronas gali būti apmokytas aproksimuoti netiesines duomenų priklausomybes. Paprasčiausia yra šuolinė aktyvacijos funkcija, kurios lygtis yra sekanti:

$$f(z) = \begin{cases} 0, & \text{jei } z < 0; \\ 1, & \text{jei } z \geq 0, \end{cases} \quad (5.2)$$

čia $z = \sum_{i=1}^n w_i x_i + b$. Ši funkcija nėra diferencijuojama taške $z = 0$, todėl gradientinio nusileidimo algoritmas (žr. 5.4 skyrių) negali būti taikomas. Ši funkcija gali būti taikoma, kai duomenys yra tiesiškai atskiriami.

Žymiai dažniau yra naudojama sigmoidinė perdavimo funkcija:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (5.3)$$

Šios funkcijos ypatybės:

- $f(z)$ yra beveik tiesinė, kai $|z|$ yra mažas;
- $f(z)$ artėja prie 0, kai $z \rightarrow -\infty$;
- $f(z)$ artėja prie 1, kai $z \rightarrow \infty$.

Ši funkcija neturi trūkio taškų, todėl perceptronui su šia aktyvacijos funkcija, gali būti taikomas gradientinio nusileidimo algoritmas. Vietoje sigmoidinės funkcijos gali būti taikoma hiperbolinio tangento funkcija, kuri duoda rezultatą iš intervalo $(-1; 1)$. Šių funkcijų problema yra ta, jog esant išėjimo reikšmei arti 0 arba 1, mokymosi greitis ženkliai sumažėja.

Šiai problemai išspręsti buvo pasiūlyta išlyginta tiesinė funkcija (angl. *rectified linear unit*), kuri yra dažnai naudojama giliojo mokymo (angl. *deep learning*) srityje. Šios funkcijos formulė yra sekanti:

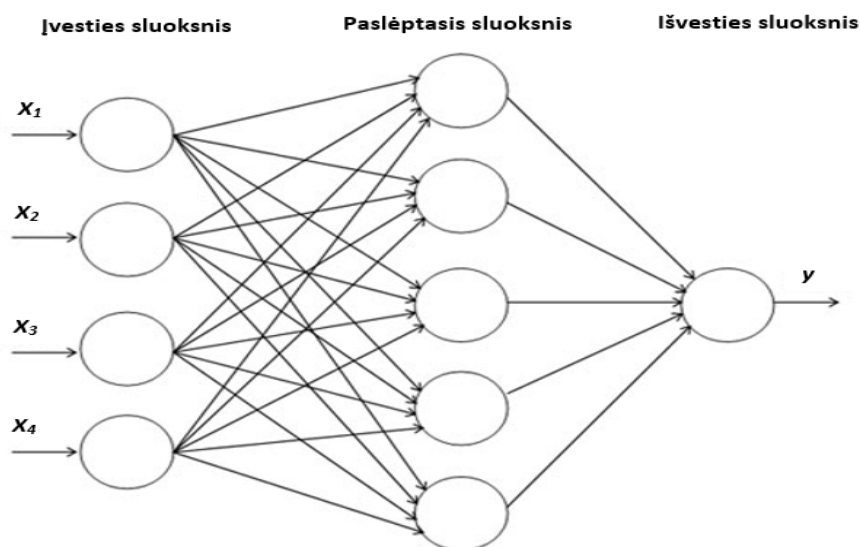
$$f(z) = \max(0; z) \quad (5.4)$$

Funkcija gražina nulį visoms neigiamoms z reikšmėms ir z , kai z yra didesnis už nulį. Ši funkcija neturi mokymosi greičio sumažėjimo problemų, kai svertinė suma z yra teigiama, tačiau su neigiamomis z reikšmėmis neuronas išvis nesimoko.

5.3. Tinklo modelis

Dirbtiniai neuronai tinklai susideda iš daugelio perceptronų. Tokio tipo modeliai vadinami daugiasluoksniais perceptronais. Modelio struktūra pavaizduota 13 pav. Modelis turi įvesties sluoksnį, vieną ar kelis paslėptuosius sluoksnius ir išvesties sluoksnį. Tokia tinklo struktūra yra vadinama tiesioginio sklidimo (angl. *feed forward*). Tokio tipo struktūroje kiekvienas neuronas yra sujungtas su kiekvienu sekančio sluoksnio neuronu. Tokios topologijos tinkle skaičiavimai vyksta nuosekliai sluoksnis po sluoksnio. Rekurentiniame tinkle sluoksnio neuronai gali būti sujungti su prieš tai buvusio sluoksnio neuronais, tačiau šiame darbe ši struktūra nebus nagrinėjama.

Dažniausiai visi vieno sluoksnio neuronai turi vienodą perdavimo funkciją [32]. Įvesties sluoksnis nėra skaičiavimo sluoksnis. Jo paskirtis yra įvesti duomenis į tinklą. Skaičiavimai atliekami paslėptuose sluoksniuose bei išvesties sluoksnyje. Išvesties sluoksnio neuronų kiekis priklauso nuo sprendžiamos problemos. Binarinės klasifikacijos bei regresijos atvejais išvesties sluoksnyje užtenka vieno neurono. Klasifikuojant kelias klases, išvesties neuronų kiekis lygus nagrinėjamų klasių kiekiui.



13 pav. Daugiasluoksnio perceptrono pavyzdys.

5.4. Modelio apmokymas

Sėkmingam regresijos ar klasifikavimo uždavinio sprendimui neuroninis tinklas turi būti tinkamai apmokytas. Tikslas yra nustatyti tokius modelio svorius w ir poslinkius b , su kuriais modelio kaštų funkcija (angl. *cost function*) būtų mažiausia. Vienas galimų kaštų funkcijos apibūrinimų yra sekantis:

$$C(\mathbf{w}, \mathbf{b}) = \frac{1}{2n} \sum_{i=1}^N (y_i - \hat{y}_i(\mathbf{w}, \mathbf{b}))^2, \quad (5.5)$$

čia y_i - tikroji išvesties reikšmė, \hat{y}_i - modelio apskaičiuojama reikšmė, N - apmokymo duomenų kiekis.

Ši kaštų funkcija vadinama kvadratine. Ji yra gana paprasta analitiniu požiūriu. Verta pastebėti, jog klasifikavimo uždaviniams spręsti labiau tinkama yra kryžminės entropijos (angl. *cross-entropy*) kaštų funkcija [32]. Akivaizdu, jog kuo mažesni skirtumai tarp apskaičiuotų ir tikrųjų išvesties reikšmių, tuo kaštų funkcija bus mažesnė. Idealiu atveju kaštų funkcija bus lygi nuliui. Optimalių parametru paieškai, su kuriais modelio kaštų bus funkcija bus mažiausia, yra naudojama klaidos sklidimo atgal algoritmas (angl. *error backpropagation*). Algoritmas yra iteracinis. Viena algoritmo iteracija vadinama epocha. Pradiniai tinklo parametrai yra inicializuojami atsitiktiniu būdu. Kiekvienoje epochoje yra apskaičiuojama klaidos funkcija $m < N$ dydžio apmokymo duomenų paketui (angl. *mini-batch*) ir atliekamas parametru atnaujinimas (žr. 5.6 formulę). Epocha baigiama, kai parametrai atnaujinami su visais N apmokymo duomenimis. Epochų skaičius yra nustatomas vartotojo. Apmokymas gali baigtis anksčiau negu nustatytas epochų skaičius, jeigu klaidos funkcija nemažėja kelias epochas iš eilės. Tokia technika vadinama ankstyvojo sustojimo (angl. *early stopping*).

$$\begin{aligned} w'_i &\leftarrow w_i - \frac{\lambda}{m} \sum_{j=1}^m \frac{\partial C_j}{\partial w_j}; \\ b'_i &\leftarrow b_i^{(\tau)} - \frac{\lambda}{m} \sum_{j=1}^m \frac{\partial C_j}{\partial b_i}, \end{aligned} \tag{5.6}$$

čia m - paketo dydis, λ - mokymosi greitis.

Parametru atnaujinimas vyksta gradientinio nusileidimo algoritmo pagalba. Šis algoritmas turi vieną parametru - mokymosi greitį. Mokymosi greitis λ yra nustatomas vartotojo. Jei λ parametras pasirinktas per mažas, klaida mažės labai lėtai, reikės daug apmokymo epochų, kad būtų pasiektas minimumo taškas. Jeigu λ per didelis - algoritmas gali "peršokti" minimumo tašką, todėl atsiras daug svyravimų. Yra sukurti algoritmai, kurie turi adaptyvų mokymosi greitį, kaip pavyzdžiui Adam [14].

5.4.1. Neuroninio tinklo parametru parinkimas

Neuroniniai tinklai yra ypač galingi, tačiau jų darbo kokybę įtakoja daugelis parametru. Pirmiausia, reikia apibrėžti tinklo architektūrą, t.y. kiek paslėptųjų sluoksnių turės tinklas, kiek neuronu bus paslėptuose sluoksniuose. Antra, reikia pasirinkti neuronu aktyvacijos funkcijas. Vieno sluoksnio neuronai dažniausiai turi vienodą aktyvacijos funkciją, tačiau skirtinguose sluoksniuose jos gali skirtis. Trečia, norint apmokyti tinklą reikia pasirinkti optimizavimo algoritmą bei tinkamą mokymosi greitį. Vieni optimizavimo algoritmai optimalius svorius bei poslinkius suranda greičiau, kitiems reikia daugiau lauko. Taip pat reiktų paminėti, jog optimalių parametru suradimui turi įtakos ir paketo dydis. Kad modelis neprisitaikytų prie nagrinėjamų duomenų ir ne persimokytų, yra naudojama regularizacija. Kaip matome neuroninis tinklas turi daug parametru, todėl tinkamų parametru parinkimas nėra trivialus ir užima daug laiko.

Rankinė paieška - šio metodo esmė yra bandymas "atspėti" tinkamus parametrus konkrečiai problemai spręsti, pasinaudojant žiniomis apie turimą problemą. Parametrai derinami atsižvelgiant į pasiektus rezultatus.

Tinklo paieška - yra nustatomi parametru intervalai tinkantys nagrinėjamai problemai spręsti. Iš šių intervalu tam tikru žingsniu pasirenkamos parametru reikšmės. Modelis yra apmokomas su visomis pasirinktomis parametru kombinacijomis. Esant dideliu kiekiu kombinacijų, visu kombinacijų įvertinimas gali užimti daug laiko.

Atsitiktinė paieška - panaši į tinklo paiešką. Yra parenkami intervalai, kuriuose gali kisti parametrai, tačiau vietoje sistemingo reikšmių paėmimo iš intervalo yra naudojamas atsitiktinis paėmimas.

Taip pat yra sukurti Bayeso tipo paieškos algoritmai, kurie panaudoja turimą eksperimentų informaciją sekantiems eksperimentams [28]. Nepaisant sukurtų automatizuotų parametrų parinkimo algoritmų, literatūroje rekomenduojama naudoti rankinę paiešką, kuri gali užtikrinti gerus, nors ne visada geriausius, rezultatus.

6. Atlikti tyrimai

6.1. Programavimo kalbos pasirinkimas

Šiame darbe buvo naudota Python programavimo kalba. Ši programavimo kalba buvo pasirinkta dėl savo paprastumo bei plataus pagalbinių bibliotekų pasirinkimo. Programos realizuojančios tam tikrą algoritmą Python kalboje reikalauja žymiai mažiau kodo eilučių negu to paties algoritmo realizacija JAVA ar C programavimo kalba. Tai palengvina kodo nagrinėjimą, taip pat lengviau aptikti programavimo klaidas. Programavimo kalba yra interpretuojama, todėl jos veikimas lėtesnis nei kompiliuojamų kalbų. Taip pat Python kalba neužtikrina tipų saugumo. Tai galėtų būti didelė problema kuriant dideles sistemas, tačiau šiame darbe ši programavimo kalbos ypatybė didesnio poveikio neturėjo. Nepaisant savo trūkumu, Python kalba yra puikus pasirinkimas tiriant signalus.

6.2. Naudotos bibliotekos

Šioje darbo dalyje bus aptartos pagrindinės bibliotekos panaudotos šiame darbe.

6.2.1. EEGTools

Ši biblioteka buvo naudojama pradinių duomenų nuskaitymui. Turimi duomenys buvo priemami EDF formatu. Europos duomenų formatas (angl. The European Data Format (EDF)) yra skirtas saugoti daug kanalų turinčius biologinius ir fizikinius signalus [3]. Nuskaičius edf rinkmeną, duomenys yra konvertuojami į žodyno tipo masyvus. Darbas su šiais masyvais yra patogus ir paprastas. Lengvai galima pasiekti kiekvieno EEG kanalo duomenis, kanalų pavadinimus, kiekvieno stebėjimo laikus, taip pat diskretizacijos dažnį. Plačiau apie biblioteką: [25].

6.2.2. PyRQA

PyRQA tai įrankis skirtas atlikti kiekybinę rekurentinių diagramų analizę. Didžiausias jo privalumas yra tas, jog jis visus skaičiavimus atlieka grafinės plokštės pagalba, pasinaudojus OpenCL karkasu. Skaičiavimai yra išlygiagretinami, todėl atliekami ypač greitai. Plačiau apie biblioteką: [24].

6.2.3. Scikit-learn

Scikit-learn yra nemokama, atvirojo kodo kompiuterio mokymosi (angl. *machine learning*) biblioteka. Joje įgyvendinti įvairūs klasifikavimo, regresijos ir klasterizacijos algoritmai. Šiame darbe ši biblioteka buvo naudojama atraminių vektorių klasifikatoriaus bei tinklo paieškos algoritmams įgyvendinti. Plačiau apie biblioteką: [23].

6.2.4. TensorFlow

Tensorflow tai populiarai atvirojo kodo biblioteka naudojama neuroniniams tinklams kurti. Ji yra sukurta Google korporacijos. Pati kompanija naudoja šia biblioteką tiek taikomajai programinei įrangai kurti, tiek įvairiems tyrimams atlikti. Tensorflow pagalba galima sukurti ypač sudėtingus neuroninių tinklų modelius. Modelio sukūrimas reikalauja daug kodo eilučių, todėl didėja tikimybė suklysti. Plačiau apie biblioteką: [6].

6.2.5. Keras

Patogesniai TensorFlow naudojimui yra sukurta Keras biblioteka, kuri leidžia aprašyti neurotinio tinklo modelį su mažesniu kodo eilučių kiekiu. Biblioteka yra nemokama, atvirojo kodo. Biblioteka įgalina greitą eksperimentavimą. Ši biblioteka - tai aukšto lygio antstatai kitoms kompiuterio mokymosi bibliotekoms. Biblioteka gali dirbti su TensorFlow, CNTK ir Theano bibliotekomis. Šiame darbe ji buvo naudojama kaip antstatai TensorFlow bibliotekai. Jos pagalba buvo sukurti ir apmokėti neuroniniai tinklai. Plačiau apie biblioteką: [11].

6.2.6. Pandas

Pandas biblioteka suteikia plačias duomenų manipuliavimo galimybes Python programavimo kalbai. Ši biblioteka yra nemokama, atvirojo kodo. Jos pagalba į programavimo aplinką buvo įkeliami apskaičiuotų rodiklių duomenys. Plačiau apie biblioteką: [17].

6.3. Rekurentinių diagramų kiekybinės analizės taikymas ligų klasifikavimui

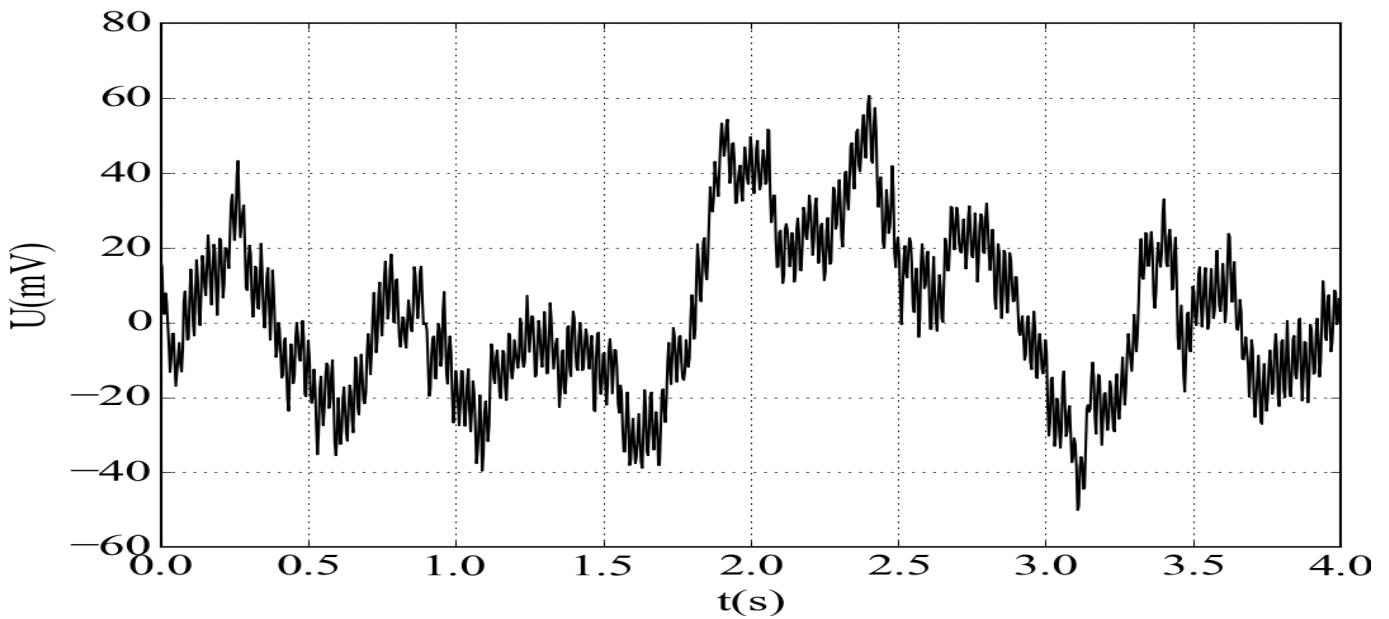
Šioje darbo dalyje bus aprašyti atlikti tyrimai pasinaudojus anksčiau aptartais metodais. Bus aptarti darbe naudoti duomenys bei jų išvalymas nuo 50 Hz elektros tinklo artefakto. Bus aprašyti atlikti tyrimai su 8, 16, 32 sekundžių ilgio EEG signalų fragmentais, taip pat EEG fragmentais, kuriuose nėra pikų, bei tyrimas su pikų sekomis. Kadangi geriausi rezultatai klasifikavimo tikslumo prasme gauti su 16 s ilgio EEG fragmentais, todėl šio tyrimo rezultatai pateikiami pirmiausia. Po to seka tyrimai su 8 ir 32 s ilgio fragmentais ir tyrimas su fragmentais be pikų. Galiausiai pristatomas tyrimas su pikų sekomis.

6.3.1. Duomenys

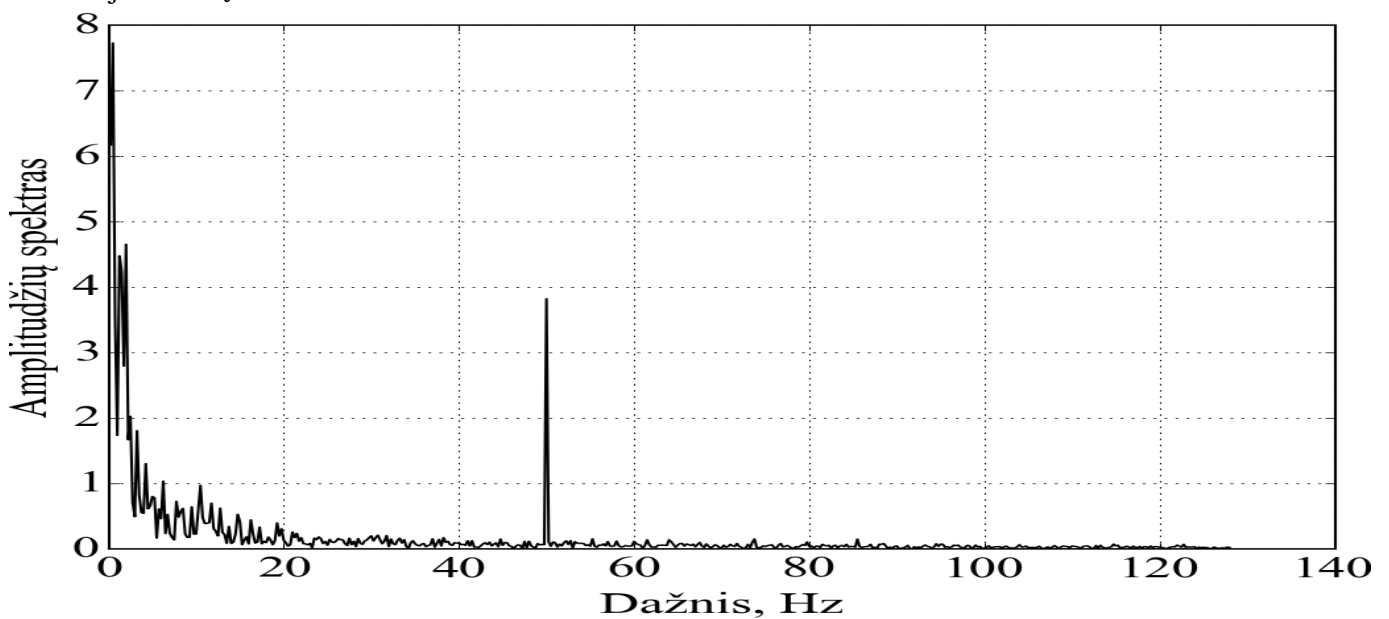
Visi šiame darbe naudoti duomenys buvo gauti iš Vaikų ligonės, VĮ Vilniaus Universiteto Santaros klinikų filialo. Tiriama pacientai sirgo Rolando epilepsija arba epilepsija, kuri buvo sukelta įvairių fizinių smegenų pažeidimų (pvz. cerebrinio paralyžiaus). Duomenys buvo prieinami EDF formatu. Kiekvienam pacientui buvo gydytojo nustatyta diagnozė ir pikų kanalas. Kiekvienas EDF failas turėjo 19 EEG kanalų. Kai kurie failai turėjo papildomus signalus, pvz. elektrokardiogramą. Failuose taip pat buvo prieinama ir kita informacija apie signalus: diskretizavimo dažnis, signalų laikas, kanalų pavadinimai. Kiekvieno paciento EEG signalai buvo skirtingo ilgio. Visų šiame darbe naudotų signalų diskretizavimo dažnis buvo 256 Hz. Reikėtų pabrėžti, jog turimuose duomenyse buvo įvairaus pobūdžio artefaktų. Šiame darbe jie nebuvo pašalinti, išskyrus 50 Hz elektros tinklo artefaktą (žr. 6.3.2 skyrių).

6.3.2. 50 Hz dažnio artefakto pašalinimas iš EEG signalo

14 pav. pavaizduotas vieno iš EEG kanalų fragmentu. Galime pastebėti, jog signalas pasižymi tam tikru "šepečiu", t. y. galime daryti prielaidą, jog signale egzistuoja tam tikra aukšto dažnio komponentė. Norėdami šią prielaidą įrodyti, mums reikės atlikti šio EEG signalo Furjė transformaciją. 15 pav. matomas tas pats EEG signalas dažniniame diapazone. Ryškus pikas egzistuoja ties 50 Hz atžyma. Tai įrodo, jog signalas yra paveiktas 50 Hz elektros tinklo artefakto. Pats



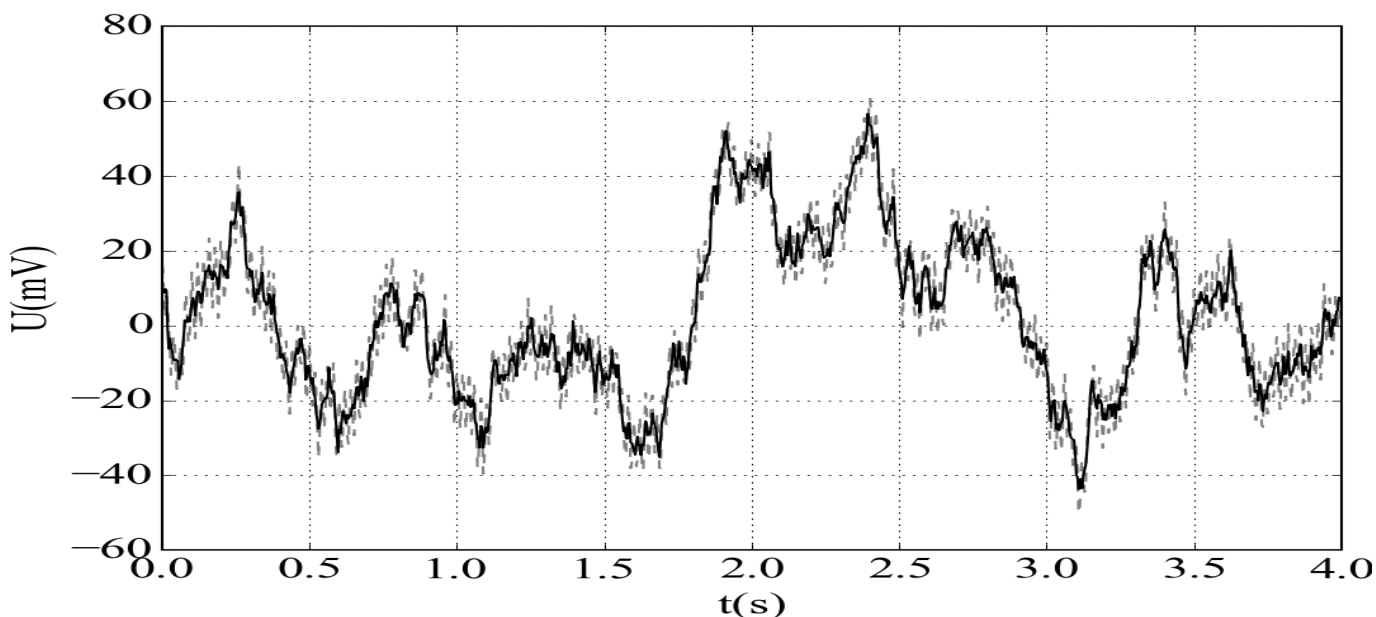
14 pav. Vieno iš elektroencefalogramos kanalų fragmentas. Aukšto dažnio osciliacijos indikuoja elektros tinklo 50 Hz artefakto būvimą. Norint griežtai įrodyti šio artefakto būvimą, reikia atlikti Furjė analizę.



15 pav. To paties EEG fragmento dažnių spektras. Smailė ties 50 Hz atžyma nurodo, jog signalė yra stipri 50 Hz komponentė.

paprasčiausias būdas pašalinti tokio tipo artefaktą yra prilyginti nuliui atitinkamus spektro koeficientus. 16 pav. buvo atlikta minėtoji transformacija. Spektro koeficientai atsakingi už 50 Hz dažnį buvo prilyginti nuliui ir iš transformuoto spektro buvo atstatytas EEG signalas laiko dimensijoje. Palyginimui tame pačiame pav. brūkšnine pilka linija atidėtas ir originalus signalas. Kaip matome, filtruotas signalas yra glodesnis už originalų.

Skaitytojas gali paprieštarauti teigdamas, jog spektriniu koeficientų prilyginimas nuliui gali ištrinti svarbią informaciją iš signalo. Žinoma, teisybės šiame teiginyje yra, tačiau kaip matome iš 15 pav., didžioji dalis spektro galios yra susikoncentravusi iki 20 Hz dažnio. Prilyginus 50 Hz



16 pav. Tas pats EEG fragmentas pašalinus 50 Hz artefaktą prilyginus nuliui atitinkamas dažnines komponentes. Palyginimui, pilka brūkšnine linija, atvaizduotas originalus signalas.

dažnį atitinkančius spektro koeficientus kaimyninių reikšmių vidurkiui (3 kaimynai iš kairės, 3 - iš dešinės), atstatyto signalo reikšmės nesiskyrė nuo tų, kurios buvo gautos "užnulinant" atitinkamus koeficientus. Todėl galima teigti, jog paprastas "užnulinimas" yra veiksmingas būdas šalinti 50 Hz artefaktą.

6.4. Tyrimas su 16 s ilgio EEG signalų fragmentais

Šioje darbo dalyje pateikiami tyrimo su 16 s ilgio EEG signalų fragmentais. Tyrimų su kitais EEG fragmentų ilgiais rezultatai pateikiami 6.5 poskyryje.

Turimi EEG duomenis EDF formatu buvo įkeliama į programavimo aplinką ir rankiniu būdu buvo pasirenkamas kanalas, kuriame gydytojas aptiko pikus. Toliau turimas EEG kanalas buvo karpomas į 4096 stebėjimus turinčius fragmentus (16 s). Toks ilgis buvo pasirinktas dėl [7] darbe naudotų tokio ilgio signalų fragmentų. Iš gautų fragmentų buvo pašalinamas 50 Hz artefaktas aukščiau aprašytu būdu, tačiau kiti artefaktai nebuvo pašalinti. Buvo panaudoti 11 pacientų sergančių Rolando epilepsija (I grupė) ir 19 pacientų su smegenų pažeidimais (II grupė) duomenys. Iš viso buvo gauta 363 kiekvienos diagnozės EEG fragmentų. Kiekvienam fragmentui buvo apskaičiuoti RD kiekybiniai rodikliai. Rekurentinių diagramų parametrai buvo sekantys: būsenų dimensija $D = 10$, vėlinimo parametras $d = 1$, slenkstinis atstumas r buvo lygus 3% maksimalaus atstumo tarp dviejų būsenų vektorių, atstumui tarp dviejų būsenų vektorių skaičiuoti buvo naudojama euklidinė metrika, minimalus įstrižos bei vertikalios linijos ilgis buvo lygus 2.

6.4.1. Statistinė RD kiekybinių rodiklių analizė

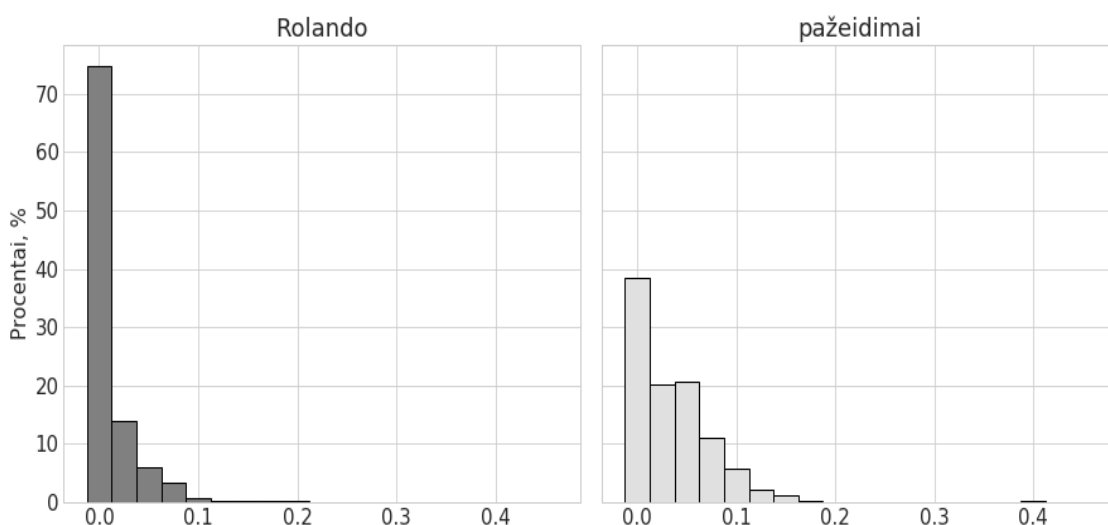
Kiekvienas efektyvus klasifikatorius turi būti paremtas atributais, kurie yra panašūs klasės viduje, tačiau skirtingi tarp klasių.

Siekiant įvertinti skirtumus tarp skirtingų grupių pacientų, prieš taikant sudėtingesnius klasifikavimo algoritmus buvo apskaičiuotos rodiklių statistinės charakteristikos, tokios kaip kvartilai, mediana bei standartinis nuokrypis. Šios charakteristikos taip pat vadinamos padėties ir sklaidos

dos charakteristikomis. Šios charakteristikos leido įvertinti duomenų pasiskirstymą abiejų grupių pacientams bei pamatyti skirtumus tarp jų. Vizualiam skirtumų įvertinimui buvo nubraižytos kiekvieno RD kiekybinio rodiklio histogramos. Kiekvienu atveju, abiejų grupių pacientų rodiklių reikšmės buvo suskirstytos į 20 intervalų ir apskaičiuotas kiekvieno intervalo dažnis.

Nubraižius histogramas buvo pastebėta tendencija, jog II grupės pacientų rodiklių histogramos yra pasislinkusios labiau į dešinę. Tą rodo ir statistinės charakteristikos, kurios visiems rodikliams buvo didesnės II grupės pacientams. Rodiklių skirstinių aprašomosios charakteristikos pateikiamos B priede. Didžiausias skirtumas buvo užfiksuotas tarp pasikartojamumo normos (RR) statistinių charakteristikų. II grupės RR statistinės charakteristikos buvo 90-476% didesnės I grupės pacientų atžvilgiu. Taigi, II grupės pacientų EEG fragmentai turėjo daugiau pasikartojančių būsenų. 17 pav. pavaizduotas RR pasiskirstymas. Rolando epilepsija sergančių pacientų grupėje, RR rodiklis susikoncentravęs pirmoje grupėje ($[0, 0.025]$ intervalas), kai smegenų pažeidimų grupėje šis rodiklis pasiskirstęs platesniame diapazone.

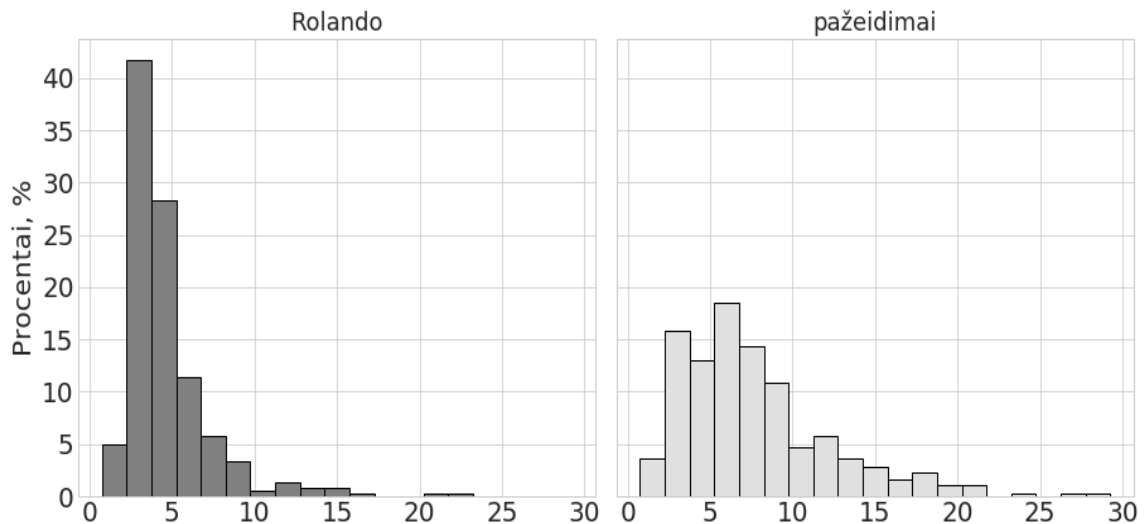
18 pav. pavaizduota vidutinio įstrižos linijos ilgio pasiskirstymo histograma. Iš jos galime matyti, jog II grupės pacientų pasiskirstymas yra platesnis, yra daugiau fragmentų su didesne vidutine įstriža linija. Tai leidžia daryti išvadą, jog II grupės pacientų fragmentai turi daugiau kvazi-periodiškumo ruožų. Tą parodo ir kitas RD kiekybinės analizės rodiklis determinizmas. Determinizmas nusako santykį tarp taškų formuojančių įstrižas linijas ir visų rekurentinės diagramos taškų (žr. 2.1.2 skyrių).



17 pav. Pasikartojamumo normos (RR) pasiskirstymas tarp skirtingų diagnozių pacientų 16 s EEG fragmentų atveju. Smegenų pažeidimus turinčių pacientų pasikartojamumo norma yra pasiskirsčiusi platesniame intervale, kai tuo tarpu Rolando epilepsija sergančių pacientų duomenys labiausiai susikoncentravę pirmoje grupėje.

19 pav. pavaizduota determinizmo pasiskirstymo histograma. Galime pastebėti, jog beveik 70% II grupės determinizmo reikšmių yra susikoncentravusios $[0.95, 1]$ intervale. I grupėje irgi dauguma rezultatų yra šiame intervale, tačiau į šį intervalą patenka tik apie 25% visų determinizmo reikšmių. I grupės pacientų determinizmo reikšmės pasiskirsčiusios platesniame intervale.

Panašus į determinizmo pasiskirstymą yra nusistovėjimo pasiskirstymas. Nusistovėjimas nusako dalį RD taškų, kurie yra vertikaliose linijose (žr 2.1.3 skyrių). Iš 20 pav. galime matyti,



18 pav. Vidutinio įstrižos linijos (L) pasiskirstymas tarp skirtingų diagnozių pacientų 16 s EEG fragmentų atveju. Smegenų pažeidimus turinčių pacientų vidutinis įstrižos linijos ilgis yra pasiskirstęs platesniame intervale, kai tuo tarpu Rolando epilepsija sergančių pacientų duomenys labiausiai susikoncentravę [0, 5] intervale.

jog antros grupės pacientų EEG fragmentuose yra daugiau atkarpų, kurios kito mažai arba nekit. Beveik 70% nusistovėjimo reikšmių yra [0.95, 1] grupėje. Rolando epilepsijos atveju, taip pat didžiausia dalis nusistovėjimo reikšmių patenka į paskutinę grupę, tačiau šių reikšmių tik yra apie 23% procentus. Tiek vienoje, tiek kitoje grupėje yra daug reikšmių, kurios patenka į pirmą grupę. Tai parodo, kad dalyje EEG fragmentų, nusistovėjusių būsenų nebuvo. I grupės pacientų atveju, žymiai daugiau nusistovėjimo reikšmių yra kituose grupėse. Taigi, galime daryti išvada, jog II grupės pacientų EEG fragmentuose yra žymiai daugiau nusistovėjimo ruožų.

Kitų rodiklių, kurių pasiskirstymo histogramos čia nėra pateiktos, tačiau jų pasiskirstymas buvo panašus į aprašytųjų, t. y. II grupės rodikliai buvo paslinkti į dešinę.

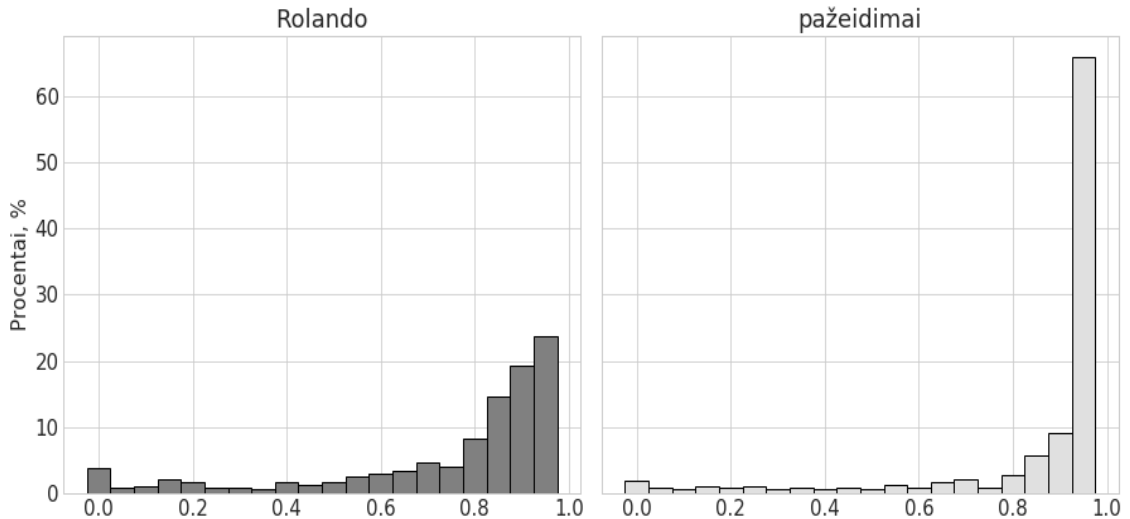
6.4.2. Klasifikavimo strategija

Įvertinus statistines rodiklių savybes dviejų grupių pacientams, šiems duomenims buvo pritaikyti atraminių vektorių bei neuroninio tinklo klasifikatoriai. Klasifikatoriams testuoti buvo naudojami apmokyme nedalyvavusių pacientų duomenys. Testavime dalyvavo 3 pacientų su diagnozuotais smegenų pažeidimais bei 2 pacientų su Rolando epilepsija duomenys. Buvo gauti 65 kiekvienos diagnozės EEG fragmentai. Klasifikatoriai buvo apmokomi su ankščiau aprašytais duomenimis (žr. 6.4 skyrių). Kadangi visi rodikliai parodė ženklus skirtumus tarp nagrinėjamų klasių, jie visi buvo naudojami klasifikatorių apmokyme.

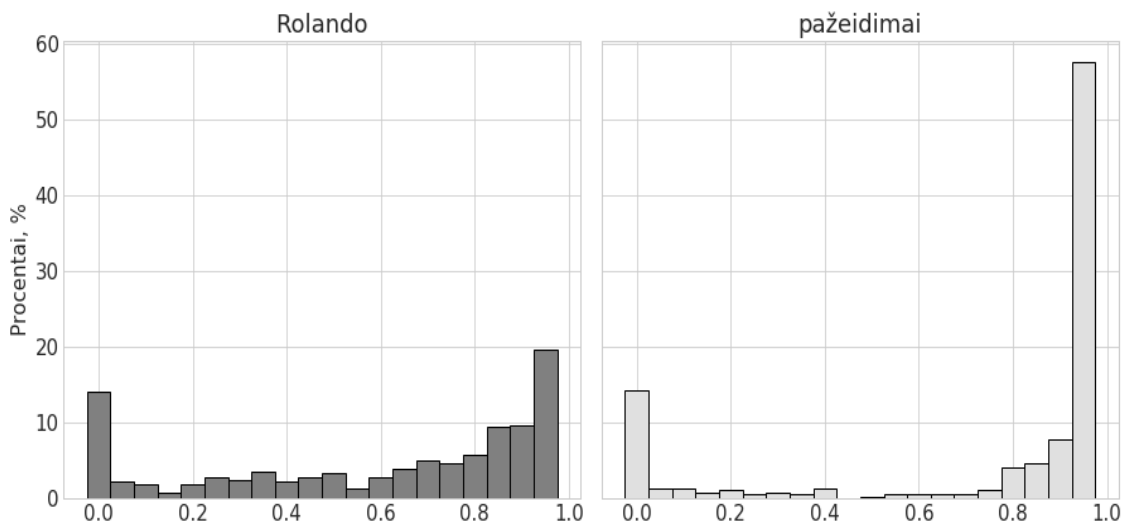
6.4.3. Atraminių vektorių klasifikatoriaus taikymas

Susijusių darbų apžvalga parodė, jog atraminių vektorių klasifikatorius gali būti naudojamas panašaus pobūdžio problemoms spręsti (žr. [7], [10] darbus).

Buvo pasirinkta [13] darbe rekomenduojama radialinių bazių branduolio funkcija. Šiai funkcijai yra reikalingi 2 parametrai: C ir γ (žr. 4.1 skyrių). Standartiniai Scikit-learn bibliotekoje



19 pav. Determinizmo (DET) pasiskirstymas tarp skirtingų diagnozių pacientų 16 s EEG fragmentų atveju. Smegenų pažeidimus turinčių pacientų determinizmo rodiklis susikongruavęs [0.95, 1] intervale, kai tuo tarpu Rolando epilepsijos pacientų atveju, jis yra pasiskirstęs platesniame intervale.



20 pav. Nusistovėjimo (LAM) pasiskirstymas tarp skirtingų diagnozių pacientų. Tie patys 20 intervalų. Smegenų pažeidimus turinčių pacientų nusistovėjimo rodiklis susikongruavęs [0.95, 1] intervale, kai tuo tarpu Rolando epilepsijos pacientų atveju, jis yra pasiskirstęs platesniame intervale. Histograma panaši į determinizmo histogramą (žr. 19 pav.)

įgyvendinto atraminių vektorių klasifikatoriaus parametrai buvo sekantys: $C = 1.0$ ir $\gamma = 1/\text{atributų skaičius}$, šiuo atveju $\gamma = 1/8$. Su šiais parametrais, klasifikavimo tikslumas su nematytų pacientų duomenimis siekė 62%.

Gerų C ir γ parametru paieškai buvo naudota tinklo paieška. Paieškai buvo naudotos C ir γ reikšmės rekomenduotos [13] darbe (žr. 4.1 skyrių). Iš visų kombinacijų, geriausia 10 sluoksnių kryžminio patikrinimo rezultatą parodė pora $C = 2^{-3}$, $\gamma = 2^{-15}$. Kryžminio patikrinimo tikslu-

mas siekė 68%. Su nematytų pacientų duomenimis šis klasifikatorius parodė 72% klasifikavimo tikslumą.

6.4.4. Neuroninio tinklo taikymas

Nagrinėjamai klasifikavimo problemai spręsti taip pat buvo panaudotas tiesioginio sklidimo neuroninis tinklas. Tinklas buvo sudarytas iš įvesties sluoksnio, vieno paslėptojo sluoksnio bei išvesties sluoksnio. Atlikus keletą eksperimentų, paaiškėjo, jog geriausius rezultatus pasiekia neuroninis tinklas su 50 paslėptųjų neuronų. Įvesties sluoksnis turėjo 8 neuronus, išvesties - 1. Paslėptajam ir išvesties sluoksniui buvo pasirinkta sigmoidinė aktyvacijos funkcija. Sviurių bei poslinkių optimizavimui buvo pasirinktas [26] darbe rekomenduojamas Adam optimizavimo algoritmas. 21 pav. pavaizduotas neuroninio tinklo klasifikavimo tikslumas. Kiekvieną kartą apmokant ir testuojant neuroninį tinklą, buvo gaunami skirtingi rezultatai. Apmokius ir ištestavus tinklą 20 kartų su tais pačiais, buvo gautas tikslumas, kuris svyravo 63-70% ribose. Vidutinis neuroninio tinklo tikslumas siekė 67%.



21 pav. Neuroninio tinklo tikslumas apmokius ir ištestavus tinklą 20 kartų. Vidutinis tinklo tikslumas siekė 67%.

6.5. Tyrimai su kitais EEG fragmentų ilgiais

Norint įvertinti EEG fragmento ilgio įtaką klasifikavimo kokybei, buvo atlikti tyrimai su 8 bei 32 sekundžių ilgio EEG fragmentais.

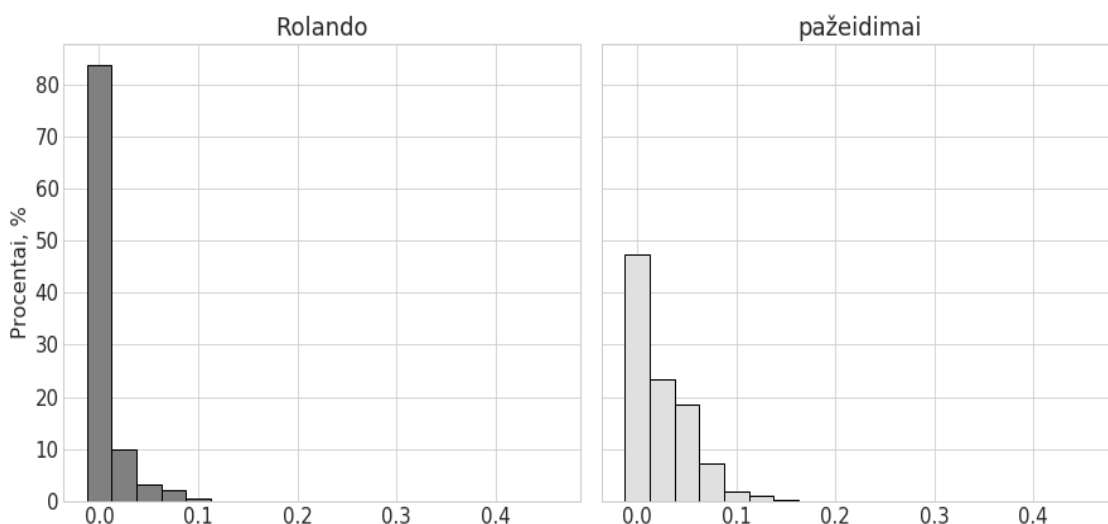
6.5.1. Duomenys ir tyrimo metodika

Tyrimo metodika buvo identiška anksčiau aprašytajai. Apskaičiuotiems RD kiekybiniais rodikliams buvo nubraižytos pasiskirstymo histogramos, įvertintoms histogramų aprašomosios charakteristikos. Tada buvo apmokyti ir ištestuoti atraminių vektorių klasifikatorius bei neuroninis tinklas. EEG fragmentų charakteristikos buvo sekanti: 8 sekundžių ilgio EEG fragmentai turėjo 2048 stebėjimus, 32 sekundžių - 8192 stebėjimus. Buvo panaudoti 19 smegenų pažeidimus

turinčių pacientų ir 11 Rolando epilepsija sergančių pacientų duomenys. Testavimui buvo naudoti duomenys gauti iš 2 pacientų sergančių Rolando epilepsija ir 3 pacientų su smegenų pažeidimais. 8 sekundžių ilgio EEG fragmentų atveju, buvo gauti 768 kiekvienos ligos fragmentai skirti algoritmų apmokymui ir 147 kiekvienos ligos fragmentais skirti testavimui. 32 sekundžių fragmentų atveju, buvo gauti 197 kiekvienos ligos fragmentai skirti apmokymui ir 36 kiekvienos ligos EEG fragmentai testavimui. Abiems atvejais, RD parametrai buvo tokie patys, kaip ir ankstesniame tyrime.

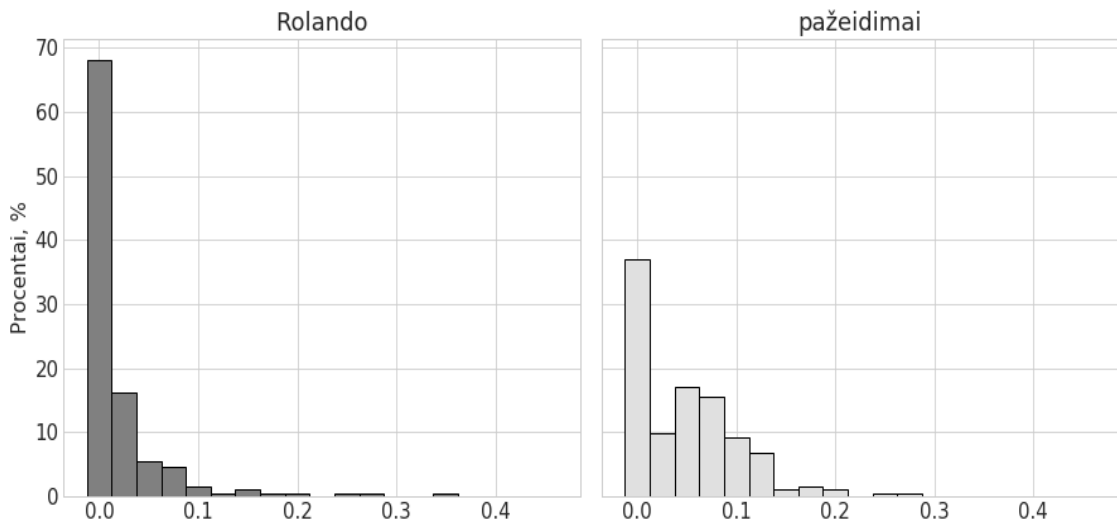
6.5.2. Statistinė analizė

Atliktos statistinės analizės rezultatai pateikiami A ir C prieduose esančiose lentelėse. Iš lentelių matome, jog kaip ir tyrimo su 16 sekundžių fragmentais atveju, II grupės kvartiliai, mediana bei standartinis nuokrypis yra didesni. Didžiausią skirtumą tarp dviejų pacientų grupių parodė pasikartojamumo normos rodiklis (RR). 8 sekundžių fragmentų atveju, kvartilių, medianos, standartinio nuokrypio skirtumai tarp I ir II grupės pacientų sudarė 34-401%, 32 sekundžių atveju, skirtumai sudarė 86-713%.



22 pav. Pasikartojamumo normos pasiskirstymas esant 8 s ilgio EEG fragmentams. Matome, kad didžioji dalis pasikartojamumo normos reikšmių Rolando epilepsijos diagnozei yra susikoncentravusios pirmame intervale. Smegenų pažeidimų grupėje pasikartojamumo normos reikšmių pasiskirstymas yra platesnis, yra daugiau reikšmių su didesne pasikartojamumo norma.

22 pav. matome pasikartojamumo normos pasiskirstymą 8 s EEG fragmentų atveju. Rolando epilepsijos EEG fragmentų RR rodiklis yra susikoncentravęs pirmoje grupėje. Smegenų pažeidimų RR reikšmės pasiskirsčiusios platesniame intervale. Tokios pat išvados galioja ir 32 sekundžių EEG fragmentams. Jų pasiskirstymas pavaizduotas 23 pav. Taigi, smegenų pažeidimus turinčios grupės EEG yra daugiau pasikartojančių būsenų. Tą patvirtina ir kiti rodikliai. Skirtingų ilgių EEG fragmentams galioja ta pati išvada, jog smegenų pažeidimų EEG fragmentai turi daugiau pasikartojančių būsenų bei daugiau kvazi-periodiškumo periodų (tą parodo vidutinis įstrižos linijos ilgis).



23 pav. Pasikartojamumo normos pasiskirstymas esant 32 s ilgio EEG fragmentams.

6.5.3. Klasifikatorių taikymas

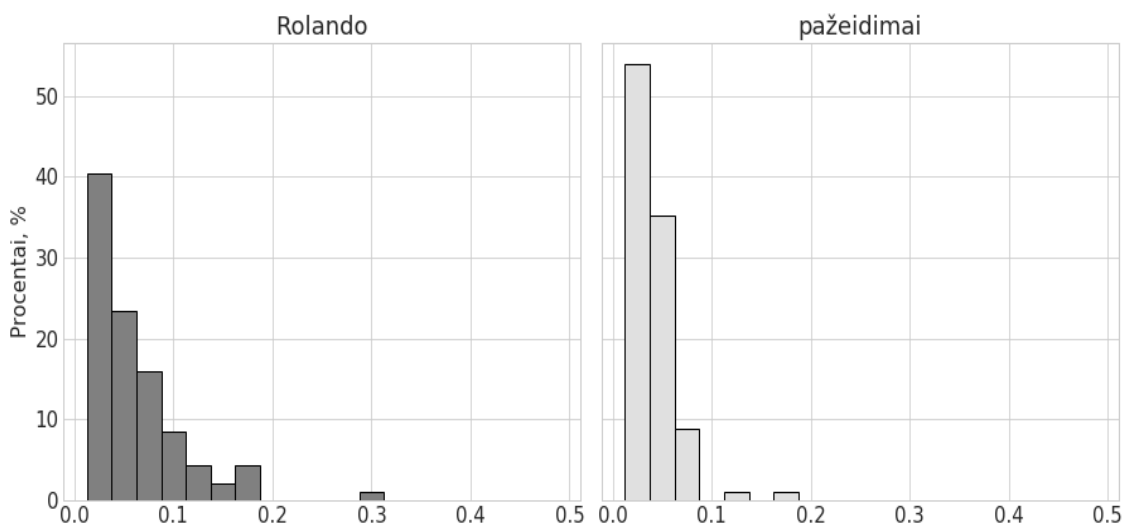
Atlikus statistinę analizę, buvo apmokyti ir ištestuoti atraminių vektorių klasifikatoriai bei neuroninis tinklas. Kaip ir ankstesniu atveju, klasifikatoriai buvo apmokomi su visais kiekybinės analizės rodikliais. 1 lentelėje pateikti klasifikavimo rezultatai testuojant klasifikatorius su nematytų pacientų duomenimis. Atraminių vektorių klasifikatorius buvo naudojamas su radialinių bazių branduolio funkcija. Gerų parametrų paieškai buvo naudota tinklo paieška (parametrų paieškos aibė identiška naudotai 16 s EEG fragmentų tyrime). Neuroninio tinklo konfiguracija buvo identiška naudotai 16 s EEG fragmentų tyrime. Kadangi neuroninio tinklo rezultatai buvo skirtingi kiekvieną kartą atliekant apmokymo ir testavimo procedūrą, todėl buvo įvertintas vidutinis neuroninis tinklo tikslumas ji apmokius ir ištestavus 20 kartų su tais pačiais duomenimis. Iš 1 lentelės matome, jog klasifikavimo rezultatai yra mažesni už atsitiktinį spėjimą ir nesiekia 50%. Prastus klasifikavimo rezultatus lėmė testavimo aibių statistinės savybės. Apmokymo aibėse apskaičiuotų duomenų padėties charakteristikos (kvartiliai ir mediana) buvo didesnės smegenų pažeidimus turinčioje grupėje, tuo tarpu testavimo aibėje padėties charakteristikos yra didesnės Rolando epilepsija sergančių pacientų grupėje. Pavyzdžiui, pasikartojamumo normos padėties charakteristikos smegenų pažeidimų grupėje buvo 32-46% mažesnės negu Rolando epilepsijos grupėje 32 s fragmentams ir 26-30% mažesnės 8 s fragmentams. 32 ir 8 sekundžių fragmentų testavimo aibių *RR* histogramas galite matyti 24 ir 25 pav. Galime matyti, jog pasiskirstymas yra priešingas apmokymo aibei. Taigi, nuo testavimo aibės parinkimo, bei EEG fragmentų kiekio priklauso klasifikavimo kokybė. Testavimo aibių visų rodiklių statistinės savybės pateiktos E ir F prieduose.

6.6. Tyrimas su kanalais, kuriuose nėra pikų

Iki šiol dirbome su EEG kanalais, kuriuose gydytojas aptiko pikus. Kaip matėme, RD kiekybiniai rodikliai parodė, jog smegenų pažeidimų grupės rodiklių aprašomosios charakteristikos yra didesnės, tačiau kai kurioms imtims šios išvados gali nebegalioti. Šioje dalyje pabandydysime atsakyti į klausimą, ar yra RD kiekybiniai rodikliai skirsis pacientų grupėms, jei vietoje kanalų, kuriuose buvo aptikti pikai, naudosime kanalus maksimaliai nutolusius nuo ligos židinio.

1 lentelė. Klasifikavimo tikslumas su nematyty pacientų duomenimis.

	Atraminių vektorių klasifikatoriaus tikslumas	Neuroninio tinklo tikslumas (vidutinis)
8 s ilgio EEG fragmentai	44%	45%
32 s ilgio EEG fragmentai	38%	43%

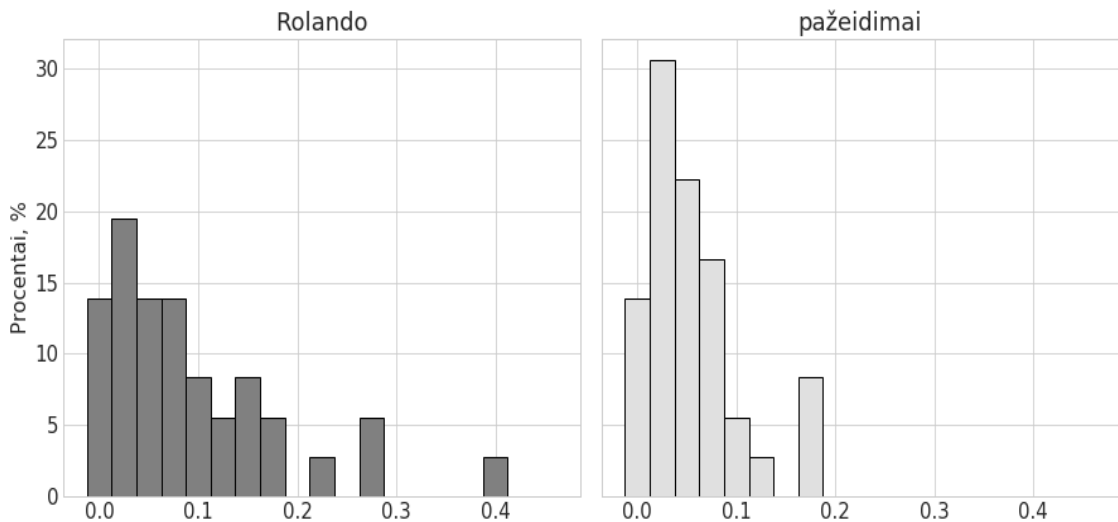


24 pav. Pasikartojamumo normos pasiskirstymas 8 s ilgio EEG fragmentų testavimo aibėje. Palygindami šią ir apmokymo aibės histogramą (žr. 22 pav.) matome, jog testavimo aibėje platesnį pasiskirstymą turi Rolando epilepsija sergančių pacientų grupė, todėl klasifikavimo rezultatai buvo prasti.

Buvo panaudoti ankstesnių tyrimų EEG duomenys, tik šiuo atveju buvo imami kanalai, kurie yra maksimaliai nutolę nuo pikų aptikimo kanalo. Kadangi geriausias klasifikavimo rezultatas buvo pasiektas su 16 sekundžių ilgio EEG fragmentais, tai ir šiuo atveju buvo pasirinktas toks EEG fragmento ilgis. Analizėje dalyvavo 396 kiekvienos ligos EEG fragmentų. Klasifikatorių testavime dalyvavo po 73 kiekvienos ligos fragmentų.

Apskaičiavus RD kiekybinius rodiklius, buvo atlikta jų statistinė analizė. Išanalizavus gautus rezultatus paaiškėjo, jog skirtumai tarp pacientų grupių yra mažesni, negu tie, kurie gauti analizuojant kanalus su pikais. Pasikartojamumo normos kvartilių skirtumai ankstesniais atvejais svyravo 173-713% intervale, šiuo atveju kvartilių skirtumai buvo 26-79% intervale. Rodiklių, susijusių su vertikaliomis linijomis, elgsena buvo priešinga anksčiau fiksuotai. Jei kanaluose su pikais atveju, II grupės pacientų vertikalių linijų rodikliai buvo didesni, tai šiuo atveju didesni yra I grupės vertikalių linijų rodikliai. Skaičiavimų rezultatai pateikiami D priede.

Klasifikatorių rezultatai buvo sekantys: atraminių vektorių klasifikatorius parodė 36% tikslumą, o neuroninis tinklas vidutiniškai parodė 46% tikslumą.



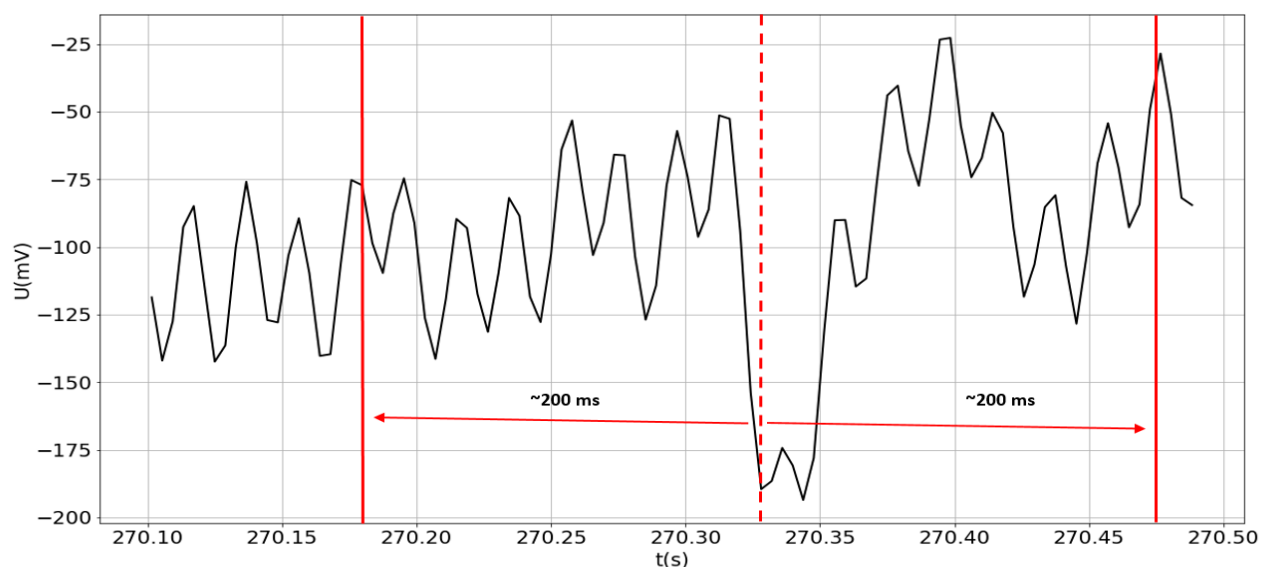
25 pav. Pasikartojamumo normos pasiskirstymas 32 s ilgio EEG fragmentų testavimo aibėje. Kaip ir 8 s ilgio fragmentų atveju, testavimo aibės savybės yra priešingos apmokymo aibės savybėms. Tai lėmė blogus klasifikavimo rezultatus.

6.7. Tyrimas su pikų sekomis

Iki šiol analizėje dalyvavo visa EEG kanale buvusi informacija. Šioje darbo dalyje koncentruosimės į pikus, kurie buvo aptikti [19] darbe aprašyto algoritmo pagalba.

Kaip ir šiame darbe, [20] darbe buvo tiriama dviejų epilepsijos tipų pacientai. Darbe aptikta tiesioginė priklausomybė tarp pikų skaičiaus, kurie dalyvauja neuroninio tinklo apmokyme ir klasifikavimo tikslumo.

Šiame darbe siekta įvertinti koks ryšys egzistuoja tarp pikų skaičiaus signale ir klasifikavimo tikslumo.



26 pav. EEG pikų iškirpimas. Punktyrinė linija - piko aptikimo vieta, nuo jos apskaičiuojamos atkarpos po apytiksliai 200 ms ir gautas EEG fragmentas iškerpamas.

2 lentelė. Klasifikavimo tikslumas su nematytų pacientų duomenimis

Pikų skaičius sekoje	SVM klasifikavimo tikslumas	Neuroninio tinklo tikslumas (vidutinis)
2	26%	24%
4	59%	49%
6	52%	51%
8	44%	49%
10	56%	50%

6.7.1. Duomenys ir tyrimo metodika

Aptikti pikai buvo iškerpami iš EEG pridėjus ir atėmus prie piko vietos po apytiksliai 200 ms (žr. 26 pav.). Taip norėta užtikrinti, jog visas pikas patektų į iškerpamą fragmentą. Kelių pikų seka buvo sujungiamą į viena signalą, kuriam buvo skaičiuojami RD kiekybiniai rodikliai.

Buvo atlikti tyrimai su 2, 4, 6, 8 ir 10 pikų viename signale. Didesnis pikų skaičius sekoje nebuvo tiriama, kadangi labai sumažėtų apmokymo ir testavimo imtys. Gautų sekų kiekis kiekvienai ligai buvo atitinkamai 762, 437, 287, 210, 170. Testavimo imtis buvo apytiksliai lygi 10% apmokymo imties dydžio. Joje nebuvo apmokyme dalyvavusių pacientų.

6.7.2. Klasifikavimo rezultatai

Statistinė rodiklių analizė neparodė kokių aiškių tendencijų tarp pikų skaičiaus signale ir skirtumų tarp nagrinėjamų grupių. Visais atvejais buvo matyti tendencija, jog smegenų pažeidimus turinčių pacientų rodiklių aprašomosios charakteristikos yra didesnės. Atraminių vektorių klasifikatoriaus ir neuroninio tinklo tikslumas pateiktas 2 lentelėje. Atraminių vektorių klasifikatorius naudojo radialinių bazių branduolio funkciją. Gerų parametrų paieškai naudota tinklo paieška. Neuroninis tinklas turėjo vieną paslėptąjį sluoksnį su 50 neuronų, vieną išvesties neuroną ir 8 įvesties neuronus. Kaip ir statistinės analizės atveju, aiškios priklausomybės tarp klasifikavimo tikslumo ir pikų skaičiaus signale nepastebėta. Verta pastebėti, jog 2 pikų signale atveju, į testavimo imtį pateko duomenys, kurių rodiklių statistinės savybės buvo priešingos apmokymo duomenų savybėms. To pasekoje pasiektas mažas klasifikavimo tikslumas. Atraminių vektorių klasifikatoriaus tikslumas buvo geresnis už neuroninio tinklo tikslumą. Dėl neuroninio tinklo parametrų gausos, naudoti tinklo parametrai galėjo būti neoptimalūs.

Išvados ir rekomendacijos

Šiame darbe buvo siekiama remiantis rekurentinių diagramų kiekybine analize klasifikuoti pacientų su smegenų pažeidimais ir Rolando epilepsija EEG signalus. Tuo tikslu buvo aptarta rekurentinių diagramų teorija, jų konstravimo principai bei parametrų parinkimo būdai. Buvo aptarti rekurentinių diagramų kiekybiniai rodikliai, kuriais remiantis galima klasifikuoti signalus. Rekurentinių diagramų kiekybinė analizė sėkmingai taikoma sveikų ir epilepsija sergančių žmonių EEG signalų klasifikavime.

Šiame darbe taip pat buvo aptarta atraminių vektorių klasifikatoriaus teorija, optimalių parametrų paieškos strategija. Taipogi aptarta neuroninių tinklų teorija. Šie klasifikatoriai naudoti atskirti Rolando epilepsija sergančių arba smegenų pažeidimus turinčių pacientų EEG pasinaudojus apskaičiuotais kiekybiniais rodikliais. Duomenys buvo gauti iš Vaikų ligoninės, VĮ Vilniaus universiteto Santaros klinikų filialo. Atlikti tyrimai parodė jog:

- 16 sekundžių EEG fragmentams atraminių vektorių klasifikatorius pasiekė 72% klasifikavimo tikslumą su apmokyte nedalyvavusių pacientų duomenimis (žr. 6.4.3 poskyrį). 20 kartų apmokius ir ištestavus neuroninį tinklą, jo klasifikavimo tikslumas su apmokyte nedalyvavusių pacientų duomenimis svyravo 63-70% intervale ir vidutiniškai siekė 67% (žr. 6.4.4 poskyrį).
- Kiekybinės analizės rodiklių, apskaičiuotų smegenų pažeidimų grupės EEG fragmentams kuriuose yra pikų, statistinės savybės (kvartilai, mediana, standartinis nuokrypis) yra didesnės negu rodiklių apskaičiuotų Rolando epilepsija sergančių pacientų EEG fragmentams. Rezultatai nesiskyrė 8, 16, ir 32 sekundžių ilgio EEG fragmentams (žr. 6.4.1, 6.5.2 poskyrius).
- Kiekybinės analizės rodiklių statistinių savybių skirtumai tarp pacientų grupių pastebimi ir EEG fragmentuose, kuriuose nėra pikų. Procentiniu požiūriu šie skirtumai yra mažesni, negu rodikliams apskaičiuotiems kanaluose su pikais. Rodiklių, susijusių su RD įstrižomis linijomis (L , RR , DET , L_{max} , $ENTR$), statistinės savybės buvo didesnės smegenų pažeidimų grupėje. Priešingai negu rodiklių apskaičiuotų EEG fragmentams su pikais atveju, čia, rodiklių susijusių su vertikaliomis linijomis (LAM , TT , V_{max}) statistinės savybės buvo didesnės Rolando epilepsijos grupėje (žr. 6.6 poskyrį).
- Dėl testinių duomenų savybių 8 ir 32 sekundžių ilgio fragmentų klasifikavimo rezultatai buvo mažesni nei 16 sekundžių ilgio fragmentų rezultatai. Atraminių vektorių klasifikatoriaus tikslumas 8 ir 32 sekundžių EEG fragmentams buvo 44% ir 38%, neuroninio tinklo klasifikavimo tikslumas atitinkamai buvo 45% ir 43%. (žr. 6.5.3 poskyrį).
- Atlikus tyrimas su įvairiu pikų kiekiu signale, nebuvo galima išskirti aiškios priklausomybės tarp pikų kiekio signale ir klasifikavimo tikslumo. Klasifikavimo tikslumas svyravo tarp 24-59% (žr. 2 lentelę). Atraminių vektorių klasifikatorius klasifikavo tiksliau negu neuroninis tinklas. Tai galėjo lemti netinkamas neuroninio tinklo parametrų parinkimas.

Ateities tyrimų gairės

- Skaičiuoti RD kiekybinius rodiklius EEG fragmentams, kuriuose yra pašalinti artefaktai (žr. 3.4 poskyrį).
- Skaičiuoti papildomus rodiklius, pvz. koreliacinę dimensiją, kurie galėtų pagerinti klasifikatorių tikslumą.
- Naudoti sudėtingesnes architektūros neuroninį tinklą.
- Klasifikatorių tikslumui įvertinti naudoti kryžminį patikrinimą, jį modifikavus tokiu būdu, jog apmokymo ir testavimo aibėse būtų skirtingų pacientų duomenys.

Padėka

Norėčiau padėkoti dokt. Andriui Vytautui Misiukui Misiūnui už aptiktų pikų duomenis, kurie buvo panaudoti pikų sekų analizėje (žr. 6.7).

Literatūros šaltiniai

- [1] Atraktorius. <https://lt.wikipedia.org/wiki/Atraktorius>.
- [2] Electrode locations of international 10-20 system for eeg (electroencephalography) recording. https://en.wikipedia.org/wiki/File:21_electrodes_of_International_10-20_system_for_EEG.svg.
- [3] European data format. <http://www.edfplus.info/>.
- [4] Kernel trick. <http://i2.wp.com/blog.hackerearth.com/wp-content/uploads/2017/02/kernel.png>.
- [5] Svm max sep hyperplane with margin. https://en.wikipedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png.
- [6] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [7] U. Rajendra Acharya, S. Vinitha Sree, Subhagata Chattopadhyay, Wenwei Yu, and Peng Chuan alvin Ang. Application of recurrence quantification analysis for the automated identification of epileptic eeg signals. *International Journal of Neural Systems*, 21(03):199–211, 2011.
- [8] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13:281–305, February 2012.
- [9] L. Chen, J. Zou, and J. Zhang. Dynamic feature extraction of epileptic eeg using recurrence quantification analysis. In *Proceedings of the 10th World Congress on Intelligent Control and Automation*, pages 5019–5022, 2012.
- [10] Shanshan Chen, Qingfang Meng, Weidong Zhou, and Xinghai Yang. *Seizure Detection in Clinical EEG Based on Multi-feature Integration and SVM*, pages 418–426. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [11] François Chollet et al. Keras. <https://github.com/keras-team/keras>, 2015.
- [12] J.J. Halford. Computerized epileptiform transient detection in the scalp electroencephalogram: obstacles to progress and the example of computerized ecg interpretation. *Clin Neurophysiol.*, 120(11):1909-15, 2009.
- [13] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.
- [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
- [15] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5):237–329, 2007.

- [16] Norbert Marwan, Niels Wessel, Udo Meyerfeldt, Alexander Schirdewan, and Jürgen Kurths. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical review E*, 66(2):026702, 2002.
- [17] Wes McKinney. *pandas: a foundational python library for data analysis and statistics*. 2010.
- [18] Tadas Meškauskas. Signalų analizė ir apdorojimas. http://www.mif.vu.lt/~meska/SAA/Tadas_Meskauskas_-_Signalu_Analize_Ir_Apdorojimas_-_Mokymo_Priemone.pdf.
- [19] Andrius Vytautas Misiukas Misiūnas. Automatinė elektroencefologramų analizė. Master's thesis, Vilniaus Universitetas, 2015.
- [20] Andrius Vytautas Misiukas Misiūnas, Tadas Meškauskas, and Rūta Samaitienė. Algorithm for automatic eeg classification according to epilepsy type: benign childhood epilepsy and structural focal epilepsy. 2017. Submitted to *Journal of Biomedical Informatics*.
- [21] Ee Ping Ng, Teik-Cheng Lim, Subhagata Chattopadhyay, and Muralidhar Bairy. Automated identification of epileptic and alcoholic eeg signals using recurrence quantification analysis. *Journal of Mechanics in Medicine and Biology*, 12(05):1240028, 2012.
- [22] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. <https://scikit-learn.org/stable/>.
- [24] T. Rawald, M. Sips, N. Marwan, and D. Dransch. *Fast Computation of Recurrences in Long Time Series*. Springer Proceedings in Mathematics and Statistics, 2014. <https://pypi.python.org/pypi/PyRQA>.
- [25] B. Redurnik. Eegtools. <https://github.com/breuderink/eegtools>.
- [26] Sebastian Ruder. An overview of gradient descent optimization algorithms. 2016. <http://arxiv.org/abs/1609.04747>.
- [27] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, 2007.
- [28] J. Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. *CoRR*, abs/1412.6980, 2012.
- [29] D. Puthankattil Subha, Paul K. Joseph, Rajendra Acharya U, and Choo Min Lim. Eeg signal analysis: A survey. *Journal of Medical Systems*, 34(2):195–212, Apr 2010.
- [30] Pang-Ning Tan, Vipin Kumar, and Michael Steinbach. *Introduction to data mining*. Boston : Pearson Addison Wesley, 1st ed edition, 2005.
- [31] Nitza Thomasson, Thomas J Hoepfner, Charles L Webber, and Joseph P Zbilut. Recurrence quantification in epileptic eegs. *Physics Letters A*, 279(1):94–101, 2001.
- [32] Antanas Verikas and Aidias Gelžinis. *Neuroniniai tinklai ir neuroniniai skaičiavimai*. Kauno technologijos universitetas, 2008.

- [33] Charles L Webber Jr and Joseph P Zbilut. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pages 26–94, 2005.
- [34] Joseph P Zbilut and Charles L Webber. Embeddings and delays as derived from quantification of recurrence plots. *Physics letters A*, 171(3):199–203, 1992.

Priedai

Dokumentą sudaro šeši priedai: A-C prieduose pateikiami RD kiekybinės analizės rodiklių aprašomosios charakteristikos skirtingo ilgio EEG fragmentams, kuriuose yra pikų, D priede pateikiamos kiekybinės analizės rodiklių aprašomosios charakteristikos EEG fragmentams be pikų, E-F prieduose pateiktos kiekybinės analizės rodiklių aprašomosios charakteristikos EEG fragmentams, kurie buvo naudojami testuoti klasifikatorių tikslumą.

A. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 8 sekundžių EEG fragmentams

3 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos 8 sekundžių EEG fragmentams

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	3,50	4,37	5,51	3,49
	II.	4,68	6,55	9,10	5,58
	Pokytis, %	34	50	65	60
$DET(\%)$	I.	50,59	81,85	91,91	31,06
	II.	82,43	95,59	98,23	27,26
	Pokytis, %	63	17	7	-12
$LAM(\%)$	I.	27,48	71,81	91,02	34,57
	II.	66,74	96,65	99,25	36,52
	Pokytis, %	143	35	9	6
$RR(\%)$	I.	0,13	0,54	1,65	3,70
	II.	0,56	2,70	5,46	4,97
	Pokytis, %	316	401	230	34
TT	I.	2,50	2,89	5,04	4,55
	II.	2,83	5,04	8,56	8,18
	Pokytis, %	13	75	70	80
$ENTR$	I.	1,64	2,01	2,34	0,67
	II.	2,10	2,57	2,98	0,74
	Pokytis, %	28	28	27	9
L_{max}	I.	18,00	69,50	265,25	409,50
	II.	69,50	476,00	1145,75	712,04
	Pokytis, %	286	585	332	74
V_{max}	I.	8,00	17,00	32,25	48,29
	II.	17,00	43,00	78,00	56,96
	Pokytis, %	113	153	142	18

B. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 16 sekundžių EEG fragmentams

4 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos 16 s ilgio fragmentams. Procentinis pokytis parodo II grupės pacientų kiekvieno RD kiekybinio rodiklio kvartilių, medianos ir standartinio nuokrypio padidėjimą (ar sumažėjimą) palyginus su I grupės pacientais. Čia I grupė - Rolando epilepsija sergantys, II - smegenų pažeidimus turintys pacientai.

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	3,83	4,62	5,99	2,28
	II.	5,17	7,45	10,37	5,15
	Pokytis, %	35	61	73	126
$DET(\%)$	I.	69,19	87,56	94,76	25,55
	II.	90,04	97,21	99,13	23,07
	Pokytis, %	30	11	5	-10
$LAM(\%)$	I.	33,70	74,31	92,41	35,03
	II.	78,06	97,71	99,60	37,28
	Pokytis, %	132	31	8	6
$RR(\%)$	I.	0,14	0,79	2,51	2,83
	II.	0,80	3,97	6,86	5,39
	Pokytis, %	476	401	173	90
TT	I.	2,43	2,89	4,41	3,41
	II.	2,97	5,59	10,80	7,46
	Pokytis, %	23	94	145	119
$ENTR$	I.	1,80	2,11	2,49	0,59
	II.	2,27	2,76	3,16	0,70
	Pokytis, %	26	31	27	19
L_{max}	I.	32,00	114,00	477,50	740,99
	II.	149,50	946,00	2398,00	1430,88
	Pokytis, %	367	730	402	93
V_{max}	I.	12,00	25,00	50,00	53,10
	II.	24,00	65,00	110,00	80,30
	Pokytis, %	100	160	120	51

C. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 32 sekundžių EEG fragmentams

5 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos 32 sekundžių EEG fragmentams

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	4,21	5,15	6,83	6,40
	II.	5,68	8,73	12,23	10,32
	Pokytis, %	35	70	79	61
$DET(\%)$	I.	81,54	92,05	96,43	25,34
	II.	94,17	98,41	99,51	16,77
	Pokytis, %	15	7	3	-34
$LAM(\%)$	I.	12,27	71,96	93,37	38,96
	II.	63,23	98,74	99,71	37,53
	Pokytis, %	415	37	7	-4
$RR(\%)$	I.	0,18	1,07	3,49	4,97
	II.	1,46	5,57	9,64	9,22
	Pokytis, %	713	419	176	86
TT	I.	2,27	3,02	4,18	7,60
	II.	2,73	7,31	13,24	15,42
	Pokytis, %	20	142	217	103
$ENTR$	I.	1,98	2,26	2,64	0,66
	II.	2,42	2,96	3,36	0,75
	Pokytis, %	22	31	27	13
L_{max}	I.	47,00	288,00	1063,00	1459,16
	II.	270,00	2119,00	4598,00	2861,28
	Pokytis, %	474	636	333	96
V_{max}	I.	9,00	33,00	60,00	161,78
	II.	19,00	93,00	160,00	202,41
	Pokytis, %	111	182	167	25

D. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės EEG fragmentams, kuriuose nėra pikų

6 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos EEG fragmentams, kuriuose nėra pikų. EEG fragmentai 16 s ilgio.

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	3,92	4,80	6,22	4,39
	II.	4,53	6,52	10,06	7,87
	Pokytis, %	16	36	62	79
$DET(\%)$	I.	74,89	90,03	95,52	25,91
	II.	84,94	94,38	98,71	19,89
	Pokytis, %	13	5	3	-23
$LAM(\%)$	I.	42,53	80,72	94,30	33,59
	II.	1,27	77,75	99,30	41,93
	Pokytis, %	-97	-4	5	25
$RR(\%)$	I.	0,21	0,99	3,00	5,14
	II.	0,38	1,25	5,02	5,87
	Pokytis, %	79	26	67	14
TT	I.	2,49	3,11	4,48	5,53
	II.	2,00	2,63	8,55	7,26
	Pokytis, %	-20	-15	91	31
$ENTR$	I.	1,85	2,16	2,53	0,65
	II.	2,09	2,58	3,11	0,66
	Pokytis, %	13	19	23	2
L_{max}	I.	44,00	165,50	679,50	823,58
	II.	45,00	213,00	2421,00	1563,73
	Pokytis, %	2	29	256	90
V_{max}	I.	14,00	28,00	54,00	71,60
	II.	2,00	20,50	82,25	98,81
	Pokytis, %	-86	-27	52	38

E. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 8 sekundžių ilgio testiniams duomenims

7 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos 8 sekundžių ilgio testiniams duomenims.

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	5,44	6,79	9,82	6,42
	II.	4,89	6,96	10,83	5,80
	Pokytis, %	-10	2	10	-10
$DET(\%)$	I.	92,51	96,02	98,38	9,90
	II.	88,49	96,21	98,220	20,22
	Pokytis, %	-4	0	0	104
$LAM(\%)$	I.	92,48	97,13	99,28	12,97
	II.	83,25	96,73	99,10	25,01
	Pokytis, %	-10	0	0	93
$RR(\%)$	I.	1,69	3,58	7,24	4,80
	II.	1,18	3,57	5,36	2,85
	Pokytis, %	-30	0	-26	-41
TT	I.	3,98	5,56	9,83	9,35
	II.	3,41	5,79	11,19	8,20
	Pokytis, %	-14	4	14	-12
$ENTR$	I.	2,32	2,63	3,07	0,58
	II.	2,18	2,65	3,19	0,72
	Pokytis, %	-6	1	4	24
L_{max}	I.	262,00	581,00	1069,50	576,05
	II.	97,00	478,00	844,50	542,51
	Pokytis, %	-63	-18	-21	-6
V_{max}	I.	33,00	52,00	89,00	68,27
	II.	28,00	50,00	86,00	59,30
	Pokytis, %	-15	-4	-3	-13

F. Rekurentinių diagramų kiekybinių rodiklių statistinės savybės 32 sekundžių ilgio testiniams duomenims

8 lentelė. RD kiekybinių rodiklių aprašomosios charakteristikos 32 sekundžių ilgio testiniams duomenims.

RD kiekybiniai rodikliai	Grupė	Q_1	Mediana	Q_3	Standartinis nuokrypis
L	I.	6,64	8,67	15,23	13,57
	II.	6,07	8,33	18,23	9,38
	Pokytis, %	-9	-4	20	-31
$DET(\%)$	I.	97,21	98,16	99,61	2,58
	II.	95,32	98,39	99,648	5,28
	Pokytis, %	-2	0	0	104
$LAM(\%)$	I.	95,50	97,64	99,37	14,22
	II.	91,11	98,79	99,81	13,37
	Pokytis, %	-5	1	0	-6
$RR(\%)$	I.	4,11	8,01	15,02	9,02
	II.	2,80	5,44	8,12	4,84
	Pokytis, %	-32	-32	-46	-46
TT	I.	4,81	6,24	15,86	20,61
	II.	4,43	7,94	23,22	14,30
	Pokytis, %	-8	27	46	-31
$ENTR$	I.	2,63	2,95	3,59	0,72
	II.	2,51	2,91	3,79	0,78
	Pokytis, %	-5	-1	6	10
L_{max}	I.	1480,75	2296,00	4833,00	2461,77
	II.	925,50	1551,00	5137,50	3004,09
	Pokytis, %	-37	-32	6	22
V_{max}	I.	68,25	97,00	230,75	306,16
	II.	65,00	95,00	239,50	183,83
	Pokytis, %	-5	-2	4	-40