



Automatinis EEG klasifikavimo algoritmas pagal pikų charakteristikas

Ataskaita už antruosius doktorantūros metus

Dokt. Andrius Misiukas Misiūnas

Email: andrius.misiukas@mif.vu.lt

EEG klasifikatorius

- Pagrindinis šių metų darbas – sukurtas automatinis EEG klasifikatorius, klasifikuojantis Rolando epilepsija ir smegenų žievės displazija sergančių pacientų EEG
- Algoritmą sudaro trys pagrindiniai žingsniai
 - 1) **Pikų detekcija**
 - 2) **Pikų charakteristikų nustatymas**
 - 3) **ANN klasifikatorius**

EEG klasifikatoriaus poreikio analizė

- Egzistuoja CNS ligų, galinčių turėti labai panašius pikus
 - Tai yra iš EEG neįmanoma pasakyti tikslios diagnozės, tik kelias galimas diagnozes
- Dažniausiai gydytojai nagrinėja pacientus su žinomomis diagnozėmis ir ligos anamnezėmis
 - Bet pasitaiko atvejų kai tenka interpretuoti EEG šių duomenų nežinant

EEG grupių parinkimas

- Atsižvelgiant į tai, kad algoritmas optimizuotas vaikų gėrybinių epilepsijų pikų paieškai, buvo atrinktos dvi pacientų grupės
 - Rolando epilepsija (vaikų gėrybinės epilepsijos forma), toliau vadinsime Grupe I
 - Struktūrinė židininė epilepsija (cerebrinis paralyžius, smegenų žievės displazija ir kt.), toliau vadinsime Grupe II

Grupių parinkimas

- Grupė I pasižymi aiškiais ir visada vienodos struktūros EEG pikais
- Grupė II gali turėti įvairių pikų
 - Dažnai skirtumas akivaizdus net ne profesionalui lyginant su Grupe I
 - Buvo tirtos tik tos II Grupės EEG, kurių negalėtų atpažinti arba abejotų patys neurologai nežinant diagnozės
 - Tai yra pikai vizualiai identiški net profesionalams

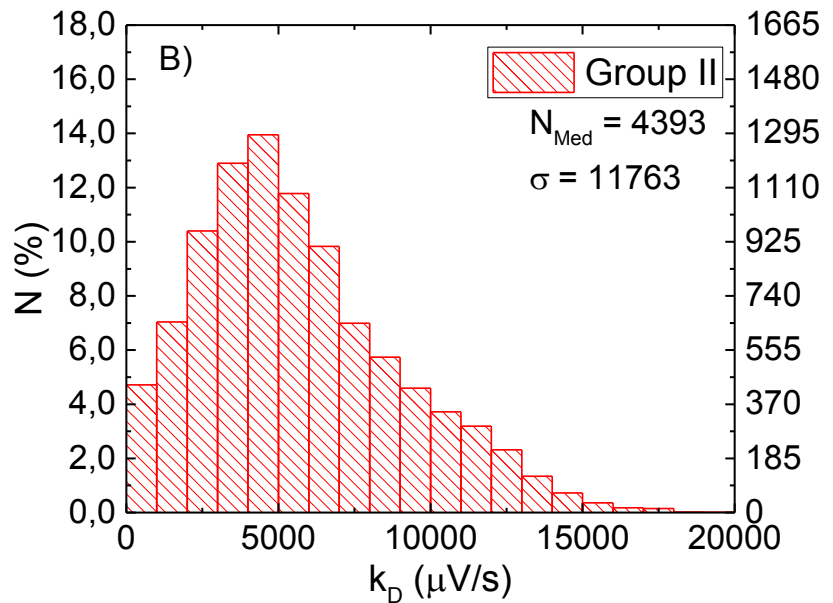
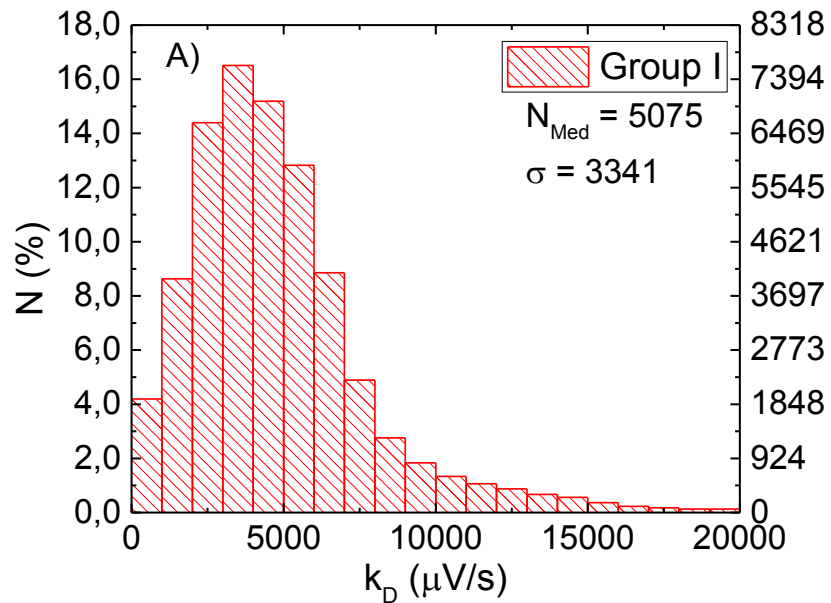
Hipotezė

- Gydytojai „iš akies“ pastebėjo, kad
 - EEG eigoje Grupės I pikai beveik nesikeičia
 - Paėmus du bet kokius pikus jie bus beveik identiški
 - Grupės II pikai išlieka panašūs, bet šiek tiek keičiasi
 - Paėmus du bet kokius pikus jie bus panašūs, bet ne visai identiški
- Todėl buvo pasitelktos jau išmatuotos EEG pikų charakteristikos ir automatinio klasifikavimo metodai šiai hipotezei patikrinti
 - Šis algoritmas – pirmųjų doktorantūros metų rezultatas

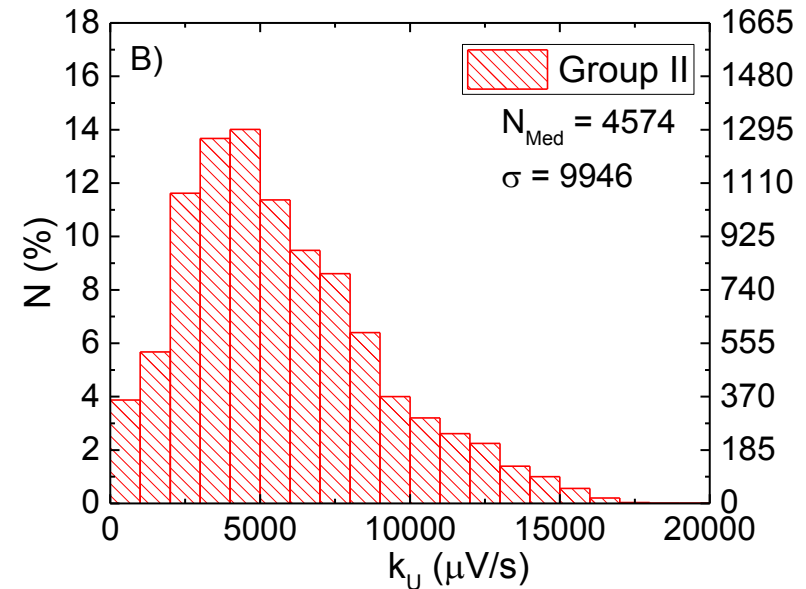
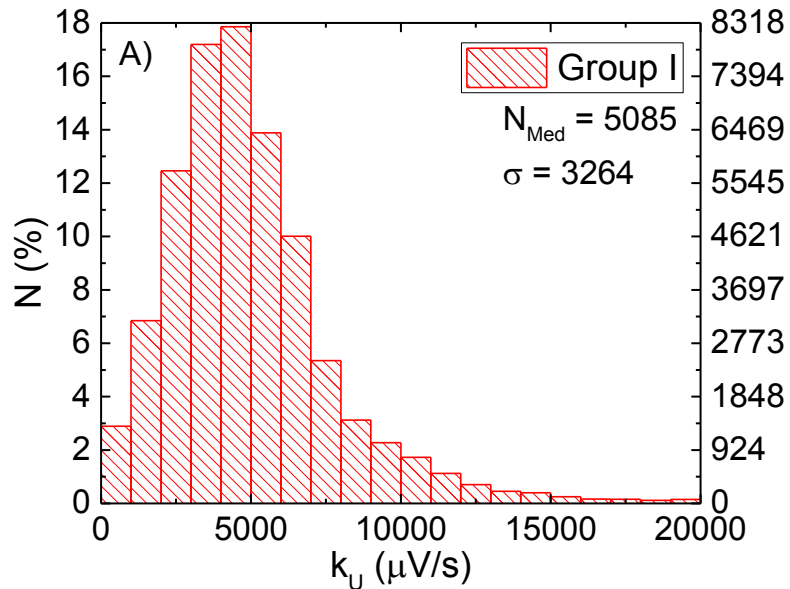
Hipotezės patikrinimas

- Turint pikų charakteristikų matavimo rezultatus šią hipotezę galima nesudėtingai patikrinti
 - Pradžiai buvo pasitelktos Origin programos statistinės analizės funkcijos
 - Pastebėta, kad charakteristikų vidurkiai skiriasi nedaug, tačiau daugiau skiriasi jų sklaida, kas patvirtina neurologų iškeltą hipotezę

Nusileidimo kampai

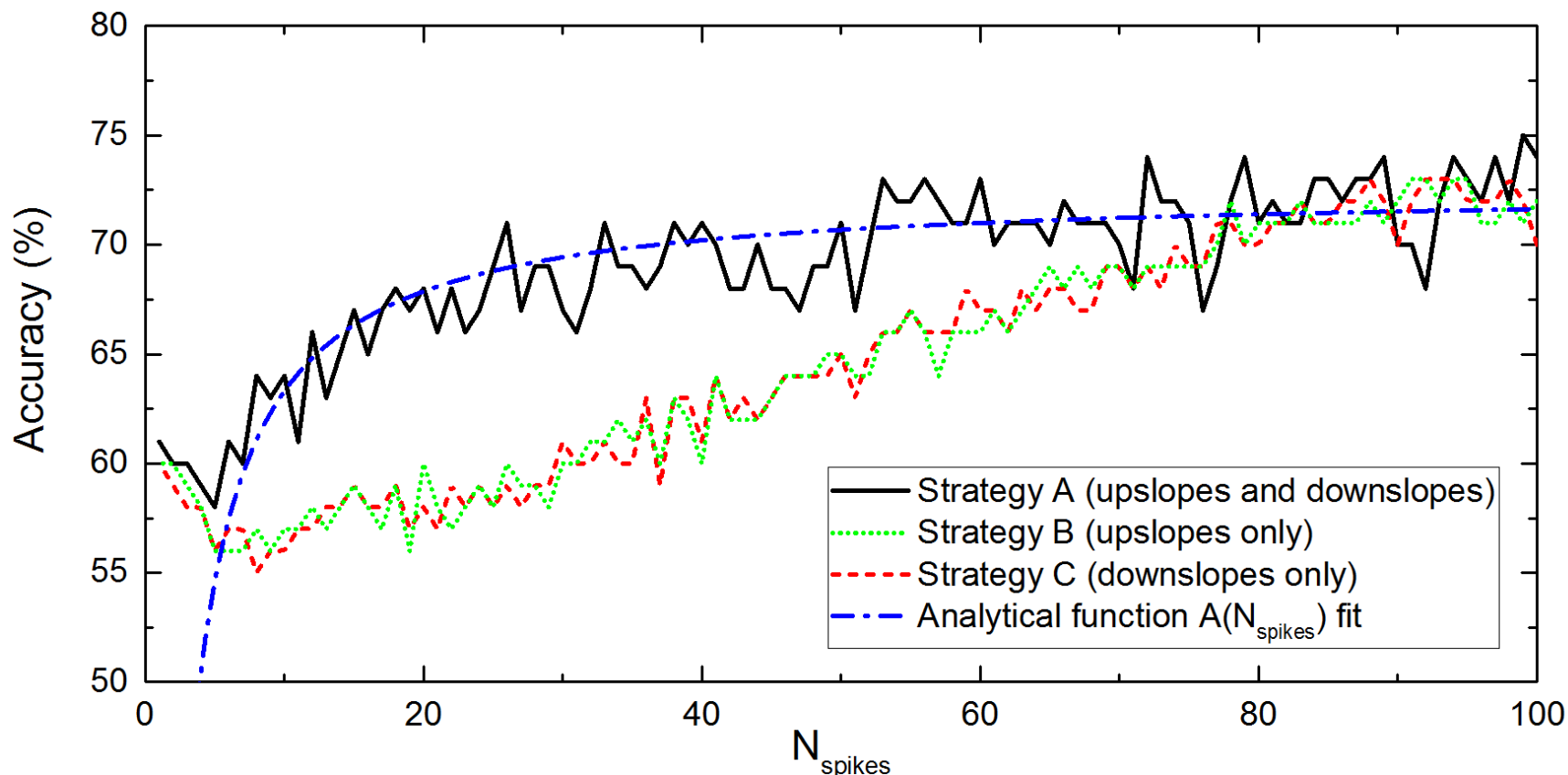


Pakilimo kampai



Klasifikatoriaus rezultatai ir duomenų kiekio įtaka jo patikimumui

- Šiuo metu patikimumas įsisotina iki 72%



Išvados ir ateities darbai

- 72% klasifikatoriaus patikimumas yra geras rezultatas, bet tikrai dar nėra tobulas
 - Tai yra **naujas rezultatas** ir algoritmas, tokio tipo duomenų dar nėra bandyta klasifikuoti automatiškai
 - Jį galima būtų tobulinti kuriant EEG artefaktų atpažinimo algoritmą
 - Bonusas: kartu pagerėtų ir pikų detekcijos algoritmo rezultatai
 - Pridėti daugiau kriterijų

DOKTORANTŪROS PLANO VYKDYMAS

Mokslinio plano vykdymas (faktiškai atlikta iki 2016 m. ataskaitos)

Planas	Terminas	Įvykdymas
2.1.1. Sudaryti EEG signalų analizės, klasifikacijos ir diagnostikos algoritmus, tame tarpe skirtus EEG pikų ir artefaktų atpažinimui.	2016.10.30	Pikų parametrų nustatymo algoritmas, pikų validacija pagal jų parametrus artefaktų atmetimui (pristatyta pernai metų ataskaitoje)
2.1.2. Parašyti šiuos algoritmus įgyvendinančias kompiuterines programas.	2016.10.30	Atlikta, iš čia gauti jų rezultatai (pristatyta pernai metų ataskaitoje)

Mokslinio plano vykdymas (faktiškai atlikta per šiuos metus)

Planas	Terminas	Įvykdymas
2.2.1. Patikrinti EEG analizei skirtų kompiuterinių ir matematinių modelių validumą.	2017.10.30	Hipotezės pradinis patikrinimas, rezultatas patvirtintas įgyvendinus EEG klasifikatorių
2.2.2. Remiantis sudarytais matematiniais ir kompiuteriniais modeliais numatyti naujus rezultatus.	2017.10.30	Panagrinėjus I ir II grupių pikus pastebėta, kad pagal šiuos požymius juos galbūt galima klasifikuoti, rezultatas patvirtintas įgyvendinus EEG klasifikatorių

Mokslinio plano vykdymas (įgyvendinta per šiuos metus, bus tęsiama toliau)

Planas	Terminas	Įvykdymas
2.3.1. Panaudojant sukurtus modelius, algoritmus ir kompiuterines programas, skaitiškai ištirti daug skirtingų pacientų (su žinomomis ligos istorijomis bei kitais duomenimis) EEG, kompiuterinius skaičiavimus atliekant tiek vienprocesorinėse, tiek daugiaprosesorinėse skaičiavimo platformose.	2018.01.30	Šiuo metu tirtos 94 EEG, bus daugiau kai atsiras daugiau II grupės pacientų
2.3.2. Nustatyti galimus sąryšius tarp kompiuterinėmis programomis apskaičiuotų EEG parametrų ir paciento diagnozės bei kitų svarbių duomenų apie pacientą.	2018.01.30	EEG klasifikatorius su 72%* patikimumu, planuojama tobulinti

* Su triukšmais, pasitaikančiais realiuose duomenyse ir nagrinėjant tik tuos atvejus, kai patys medikai vien iš EEG nepasakytų tikslios daignozės

Kitų metų mokslinis planas

Planas	Terminas	Įvykdymas
2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas:		
2.4.1. Optimizuoti kompiuterinę EEG analizės metodiką, rezultatų pateikimo būdus bei jų pritaikymą (interaktyvioje) EEG diagnostikoje.	2018.10.30	Bus atliekama šiemet
2.4.2. Pasiūlyti naujus automatizuotos ar interaktyvios EEG analizės algoritmus, tame tarpe EEG pikų ir artefaktų paieškos algoritmus.	2018.10.30	Jau pasiūlyti anksčiau aptarti algoritmai, bus kuriami nauji

Konferencijos

- Lietuvos Matematikų Draugijos konferencija, suplanuotas sudalyvavimas 2017-2019 metais
 - Dalyvauta LMD 56 2015 m. birželio mėn. 16-17d. Su pranešimu „On the implementation and improvement of automatic EEG spike detection algorithm“
 - Dalyvauta LMD 57 2016 m. birželio mėn. 20-21d. Su pranešimu „Derivative parameters of electroencephalograms and their measurement methods“

Straipsniai

- Publikuota (recenzuojami, ne periodiniai, plane nenumatyti)
 - A.V. Misiukas Misiūnas, T. Meškauskas, A. Juozapavičius „On the implementation and improvement of automatic EEG spike detection algorithm“ *Proc. of the Lithuanian Mathematical Society, Ser. A, 56, 2015, 60–65*
 - A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė „Derivative parameters of electroencephalograms and their measurement methods“ *Proc. of the Lithuanian Mathematical Society, Ser. A, 57, 2016, 47–52*

Straipsniai

- Ruošiama publikacijai, yra pilnas tekstas (pagal planą jau turėjo būti publikuotas) (periodinis, recenzuojamas)
 - A.V. Misiukas Misiūnas, T. Meškauskas, R. Samaitienė „Algorithm for automatic classification of benign childhood epilepsy and structural focal epilepsy patient EEGs“, Health Informatics Journal (Q2 žurnalas)
 - Paaiškinimas dėl vėlavimo: tam, kad gauti šiuos rezultatus reikėjo atlikti tuos mokslinio plano darbus, kurie buvo numatyti pusmečiu vėliau nuo dabartinės datos

Egzaminai

- Išlaikyti egzaminai
 - Lygiagretieji ir paskirstytieji skaičiavimai
 - Prof. J. Žilinskas, Prof. Habil. Dr. G. Dzemyda, Dr. V. Medvedev (įvertinimas 10)
 - Skaitiniai metodai
 - Prof. F. Ivanauskas, Prof. T. Meškauskas, Doc. P. Katauskis (įvertinimas 10)
 - Dirbtiniai neuroniniai tinklai
 - Prof. Š. Raudys, Dr. V. Dičiūnas, Dr. R. Kybartas (įvertinimas 9)

Egzaminai

- Kitų egzaminų situacija (pagal planą jau turėjo būti išlaikyti)
 - Duomenų vizualizavimo metodai
 - Padaryta praktinė užduotis, atliekami jos pataisymai, kuriuos nurodė prof. habil. dr. G. Dzemyda

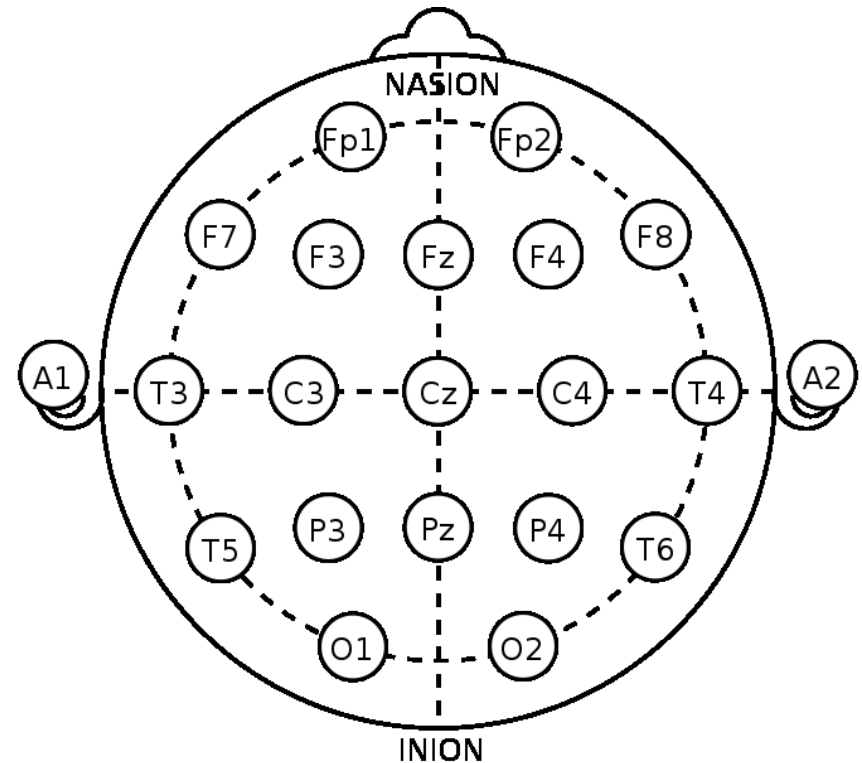
Apibendrinimas

- Šiais metais gauta naujų mokslinių rezultatų, kurie tikrai bus publikuoti
- Nors formaliai atsilieikama nuo kai kurių plano punktų vykdymo, realiai yra vykdomi kiti punktai
 - Sudalyvauta LMD konferencijoje anksčiau, negu numatyta ir jau du kartus vietoje vieno
 - Publikuoti du neplaniniai straipsniai LMD darbuose
 - Mokslinis planas vykdomas į priekį tam, kad parengti geriausios įmanomos kokybės publikacijas tarptautiniuose periodiniuose recenzuojamuose žurnaluose

AČIŪ UŽ DĖMESĮ!

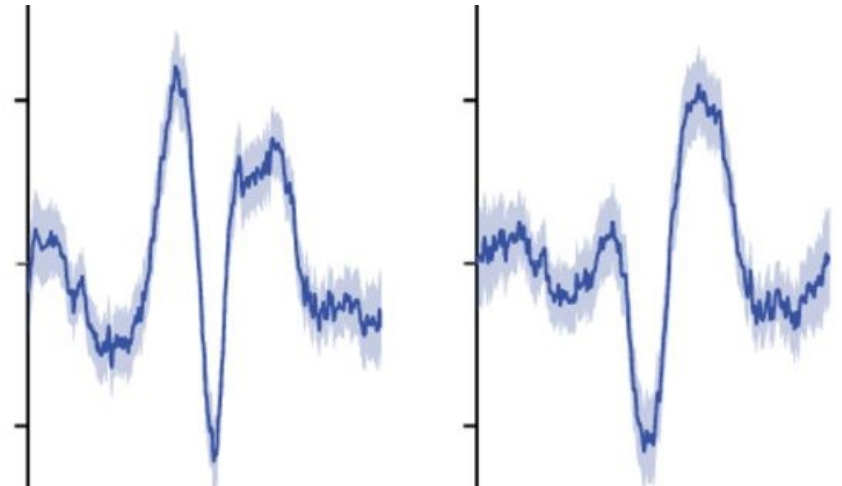
EEG

- Naudojamos 10-20 sistemos elektroencefalogramos (EEG)
- Naudojami Santariškių vaikų ligoninės pateikti duomenys



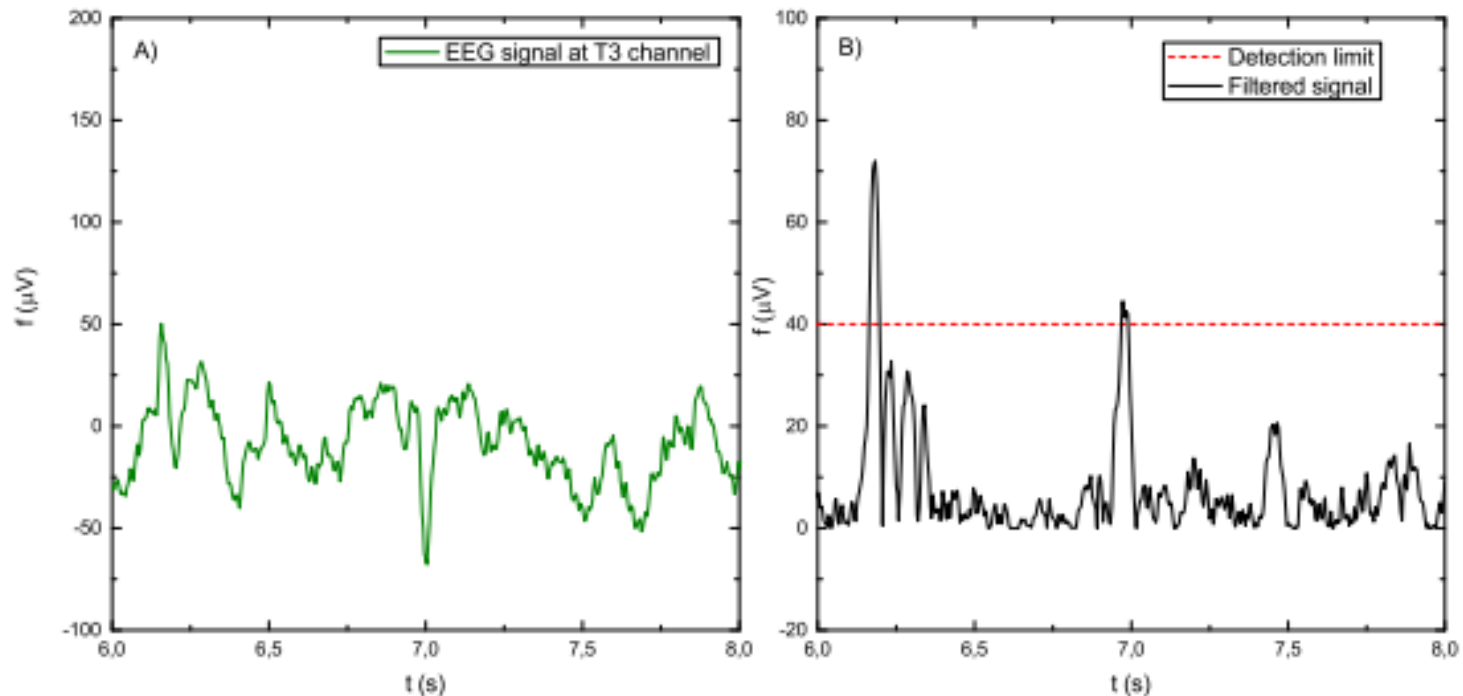
EEG pikai

- Turime EEG pikų paieškos algoritmą
- EEG pikais pasižymi įvairios epilepsijos formos ir kai kurios kitos CNS ligos



Pikų paieškos algoritmas

- Pagrįstas matematinė morfologija
- Išfiltruoja normalų smegenų aktyvumą paliekant nenormalų, pvz. pikus

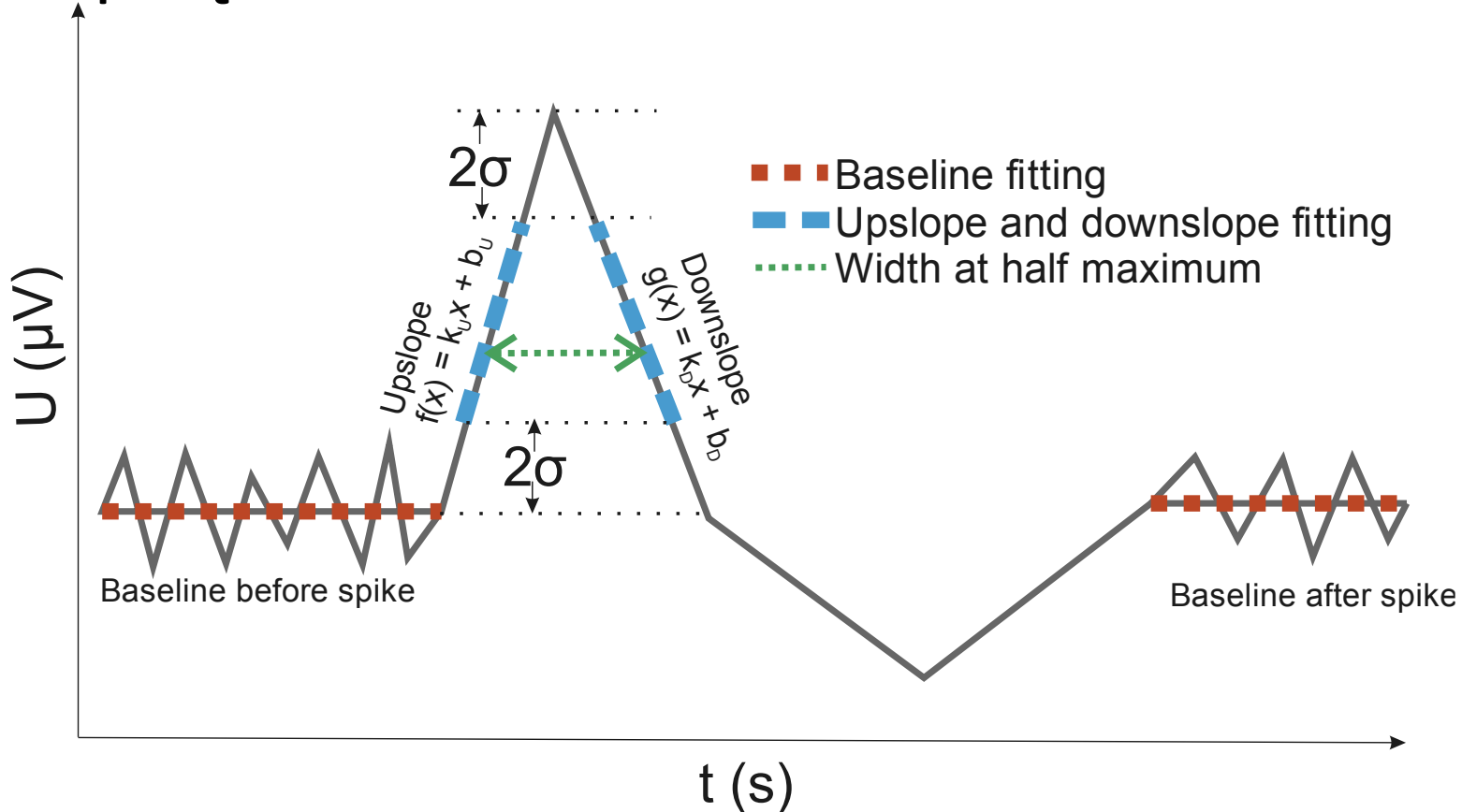


Pikų paieškos algoritmo trūkumai ir jų taisymas

- Pradinis algoritmas nėra tobulas
 - Pasitaiko netikrų teisingų detekcijų (aptinkami
 - Pasitaiko netikrų neteisingų detekcijų (neaptinkami pikai)
- Patikimumas pagerintas praeitais metais apskaičiuojant pikų charakteristikas ir jas validuojant

EEG parametrai

- EEG parametrai, naudojami klasifikacijoje ir pikų validavime



Klasifikavimo uždavinys

- Turime klasikinį binarinį klasifikavimo uždavinį
 - Standartinis sprendimas – SVM
- Šiuo atveju netiko, patikimumas iki 52%
 - Paėmus vieno piko pakilimo ir nusileidimo kampus neįmanoma pasakyti kurioje imtyje jis yra
 - Paėmus daugiau parametrų Matlab ir Python bibliotekos nerado tinkamos hiperplokštumos
 - Tikėtina, kad parametrų debesys persidengia ir aukštesnėse dimencijose

Dirbtinių neuroninių tinklų pritaikymas

- Panagrinęjus teoriją pritaikytas vieno sluoksnio perceptrono ANN
 - Įvesties sluoksnio neuronų – tiek pat kiek parametru
 - 20 paslėptų neuronų (parinkta eksperimentiškai)
 - 1 išvesties neuronas, nes turime tik vieną išvesties parametru – EEG grupę

Reikšmingų parametru parinkimas

- Buvo išbandyti ANN klasifikatoriai su visais turimais EEG pikų parametrais
 - Pakilimo kampas, nusileidimo kampas, plotis pusaukštyje, bazinė linija
- Nustatyta, kad didžiausią reikšmę turi pakilimo ir nusileidimo kampai
 - Įtraukiant ar pašalinant kitus parametru klasifikavimo patikimumas negerėjo, todėl kiti parametrai buvo išmesti iš klasifikatoriaus

Apmokymo strategija

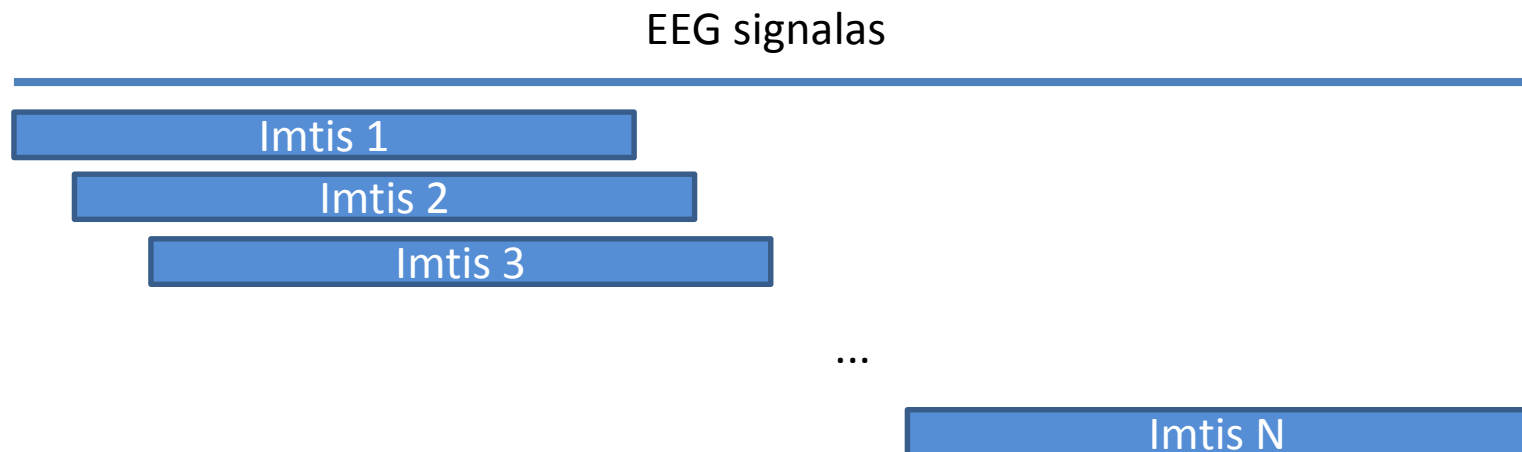
- Turėtos dviejų rūšių EEG – su medikų iškarpytais artefaktais ir visos
 - Iškarpytos naudotos apmokymui, visos – testavimui
- Keli pacientai turėjo po daugiau negu vieną EEG
 - Tokiais atvejais visos jo EEG keliavo arba į testavimo arba į mokymo imtį

Apmokymo strategija

- Iš abiejų imčių mokymo imtis sudarė apie 1/3 visų duomenų
 - 62 I Grupės EEG (21 – mokymui, 41 – testavimui)
 - 32 II grupės EEG (10 – mokymui, 22 – testavimui)

ANN apsaugojimas nuo *overfitting*

- Esant mažai duomenų ANN gali išmokti ne kaip juos apibendrinti, o jų anomalijas
 - Naudota validacija
 - Apmokymas stabdomas kai rezultatai ima prastėti su validacijos sub imtino
 - Naudotas duomenų imčių perdengimas
 - Dirbtinai padidinamas apmokymo imties dydis



Duomenų kiekio įtaka klasifikatoriaus patikimumui

- Kuo daugiau duomenų turime – tuo didesnis klasifikatoriaus patikimumas
 - Įsisotinimas pasiektas su 50-100 pikų duomenimis
 - EEG klasifikatorius gali būti naudojamas ir su mažiau duomenų, bet su mažesniu patikimumu
 - Galima naudoti ir tik pakilimo arba tik nusileidimo kampus, bet tada reikia turėti dvigubai daugiau EEG pikų