

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 531

**Razvrstavanje elektroencefalografskih
signala temeljeno na izlučivanju značajki**

Lea Suć

Zagreb, lipanj 2013.

Posebno zahvaljujem mentoru prof. dr. sc. Nikoli Bogunoviću i asistentu dr. sc. Alanu Joviću na pomoći i savjetima oko izrade rada. Zahvaljujem i svojoj obitelji, a najviše sestri, na podršci.

Sadržaj

1.	Uvod	1
2.	Snimanje i pohrana elektroencefalograma	5
2.1.	Ljudski mozak, prijenos akcijskog potencijala	5
2.2.	Elektroencefalografija	7
2.3.	Mjerenje EEG-a (međunarodni sustav 10-20)	8
2.4.	EDF format	10
3.	Izlučivanje značajki iz EEG-a	12
3.1.	Koncept dinamičkog sustava	12
3.1.1.	Nelinearni dinamički sustavi	13
3.2.	Nelinearna analiza vremenskih nizova	16
3.1.1.	Rekonstrukcija dinamike iz promatranja – ugrađivanje (engl. <i>embedding</i>)	16
3.1.2.	Karakterizacija rekonstruiranog atraktora	19
3.1.3.	Provjera valjanosti testiranjem sa zamjenskim podacima	19
3.2.	Pregled implementiranih EEG značajki	20
3.3.	Opis značajki korištenih u analizi epilepsije	23
3.3.1.	Srednja vrijednost sirovih signala (engl. <i>mean</i>)	23
3.3.2.	Hurstov eksponent (engl. <i>Hurst exponent</i>)	23
3.3.3.	Korelacijska dimenzija (engl. <i>correlation dimension, D_2</i>)	24
3.3.4.	Najveći Lyapunovljev eksponent (engl. <i>largest Lyapunov exponent</i>)	25
3.3.5.	Entropija uzorka (engl. <i>sample entropy, SampEn</i>)	25
3.3.6.	Mjera središnje težnje (engl. <i>central tendency measure, CTM</i>)	26
3.3.7.	Složenost Lempel-Ziv (engl. <i>Lempel-Ziv (LZ) complexity</i>)	26
3.3.8.	Pogreška nelinearnog predviđanja (engl. <i>nonlinear prediction, NLP</i>)	28
3.3.9.	Zajednička dimenzija (engl. <i>mutual dimension, D_m</i>)	29
3.3.10.	Značajke cross-recurrence plota (CRP)	29
4.	Opis razvijenog radnog okvira i primjeri korištenja	31
4.1.	Mogućnosti i primjeri korištenja EEGFramea	32
4.1.1.	Opcije za prikaz signala	33
4.1.2.	Funkcionalnost za izlučivanje značajki	36

4.1.3. Dodatne opcije	43
5. Analiza epilepsije.....	46
5.1. Dinamička priroda epileptičkih napadaja	46
5.2. Detekcija i predviđanje napadaja	47
5.3. Primjena EEGFramea za analizu epilepsije	48
5.3.1. Cilj i opis provedbe analize	48
5.3.2. Analiza podataka	50
5.3.3. Analiza rezultata i mogućnosti poboljšanja	55
6. Zaključak.....	57
7. Literatura.....	58
Razvrstavanje elektroencefalografskih signala temeljeno na izlučivanju značajki	62
Sažetak	62
Ključne riječi	62
Feature extraction based classification of electroencephalographic signals.....	63
Abstract.....	63
Keywords	63

1. Uvod

Elektroencefalografija je neurofiziološka mjerna tehnika kojom se pomoću elektroda na glavi registrira moždana električna aktivnost i rezultira zapisom koji se naziva elektroencefalogram (EEG). Električna aktivnost mozga rezultat je međudjelovanja živčanih stanica (neurona), kojih u ljudskom mozgu ima oko 10^{11} , pri čemu je svaki povezan s oko 10^4 drugih, a prijenos informacija među njima elektrokemijske je prirode. Električni potencijal koji stvara pojedinačni neuron ima izrazito malu amplitudu, no sinkronizirano djelovanje čitave mreže neurona može se uspješno mjeriti na ovaj način.

U oscilacijama EEG-a postoje važne informacije o tome kako mozak funkcionira. Za različita stanja svijesti (npr. budnost, san) i za različita patološka stanja (npr. epilepsija, moždani udar) postoje karakteristični EEG signali, zbog čega je važna primjena EEG-a u dijagnostici bolesti, ali je koristan i kod potresa mozga, gubitka pamćenja ili poremećaja spavanja, kao i ispitivanja kako funkcionira mozak zdravih osoba pri određenim aktivnostima, kao što su učenje ili meditacija. Zbog velike količine informacija prikupljenih pomoću elektroda, analiza moždanih signala vrlo je zahtjevna i važno je odrediti koji su signali bitni za analizu.

Posebno zanimljivi EEG signali karakteristični su za epilepsiju, koja pogađa 1 % svjetske populacije i predstavlja jedan od najčešćih neuroloških poremećaja, odmah nakon moždanog udara. Može se definirati kao poremećaj mozga karakteriziran trajnom predispozicijom za generiranje epileptičkih napadaja, koji mogu biti popraćeni poremećajem ili gubitkom svijesti, psihičkim, autonomnim ili osjetilnim simptomima, ili motornim fenomenima. Kod mnogih pacijenata s epilepsijom napadaji su dobro kontrolirani trenutno dostupnim antiepilepticima. Međutim, unatoč pažljivo optimiziranim terapijama lijekovima, oko 30 % pacijenata i dalje ima napadaje [1]. Kod pacijenata koji pate od napadaja koji potječu iz ograničenog područja mozga i ne mogu se uspješno kontrolirati antiepilepticima, neurokirurgija može imati 60 % - 70 % šanse za donošenje dugoročnog prestanka simptoma. Za epileptičare koji ne postižu potpunu kontrolu napadaja trenutno dostupnim terapijama,

postoji snažna potreba za novim ljekovitim tretmanima. S obzirom na činjenicu da iznenadne, nepredvidive pojave napadaja predstavljaju jedan od najviše ograničavajućih aspekata bolesti, postupak koji bi mogao predvidjeti pojavu napadaja mogao bi značajno unaprijediti terapijske mogućnosti i poboljšati kvalitetu života ljudi koji boluju od epilepsije. Zbog bliske veze s fiziološkim i patofiziološkim funkcijama mozga EEG u epileptologiji ima važnu ulogu za kliničku praksu. Za otkrivanje epilepsije potreban je EEG zapis trajanja jedan tjedan. Pregledavanje cijele duljine takvih zapisa za dobro uvježbanog neurologa zamorno je i dugotrajno [2]. Ključno je pitanje mogu li se iz kontinuiranog EEG-a izlučiti karakteristične značajke koje mogu predvidjeti nadolazeći napadaj. Kad bi se iz dinamičkih promjena u EEG-u pouzdano mogle predvidjeti pojave napadaja epileptičara, mogao bi se izmisliti potpuno automatizirani pouzdani sustav za detekciju napadaja, čime bi se uvelike smanjilo i vrijeme potrebno neurologu za pregled opsežnih EEG podataka. Kad bi takav sustav postojao, preventivne strategije liječenja (npr. dugoročno liječenje antiepilepticima, koji mogu uzrokovati kognitivne ili druge neurološke nedostatke) mogle bi se zamijeniti terapijom na zahtjev, ili bi se električnim ili drugim stimulacijama moglo pokušati vratiti dinamiku mozga u stanje iz kojeg se više neće razvijati napadaji [1]. No, to je idealistički cilj, a za početak bi bilo više nego dovoljno kad bi se iz EEG zapisa moglo točno uočiti i identificirati pojave takvih napadaja. Javlja se potreba za rafiniranim tehnikama analize koje omogućuju izlučivanje relevantnih informacija iz dugoročnih EEG zapisa.

Kako bi se omogućila poboljšana karakterizacija dinamike EEG-a, brojne tehnike linearne analize razvijene su tijekom zadnjih nekoliko desetljeća, koje se danas široko koriste u kliničke svrhe [1]. Takve metode temelje se prvenstveno na pretpostavci da EEG generira vrlo složen linearan sustav, što rezultira karakterističnim svojstvima signala kao što su nestacionarnost ili nepredvidivost. S druge strane, EEG signali mogu se tumačiti kao izlaz determinističkog sustava relativno niske složenosti, ali koji sadrži visoko nelinearne elemente [3].

Iako su linearne tehnike dale važan doprinos za razumijevanje fizioloških i patofizioloških uvjeta u mozgu, one pružaju samo ograničenu informaciju o dinamičkim aspektima EEG-a. Njima se ne može u potpunosti okarakterizirati komplicirano, očito

nepravilno ponašanje mozga kao složenog dinamičkog sustava. Nelinearnost je u mozgu prisutna već na staničnoj razini, budući da je dinamičko ponašanje pojedinih neurona regulirano integracijom, pragom i fenomenom zasićenja. Unatoč tim poznatim fiziološkim činjenicama, u više navrata se tvrdilo da možda neće biti valjano za očekivati da se velika mreža takvih nelinearnih elemenata također ponaša na nelinearan način [1]. Međutim, za razliku od normalne aktivnosti, epileptički napadaji su vrlo nelinearni fenomeni. Posljednja dva desetljeća pokazala su da nelinearna analiza vremenskih nizova (engl. *nonlinear time series analysis*) izvedena iz nelinearne dinamike (teorija kaosa) omogućuje poboljšanu karakterizaciju epileptičkih stanja mozga, pomaže u dobivanju dubljeg uvida u prostornu i vremensku dinamiku epileptičkih procesa i značajna je u razvoju upozorenja i preventivnih tehnika [1].

Općenito, deterministički kaos nudi izvrsno objašnjenje za naizgled nepravilno ponašanje mozga koje se očituje u EEG-u. Nedavni razvoj u fizikalno-matematičkom okviru teorije nelinearne dinamike pruža nove koncepte i snažne algoritme za analizu vremenskih nizova, čime se otvara niz novih perspektiva za proučavanje normalne i poremećene funkcije mozga [4]. Nelinearna se analiza vremenskih nizova uspješno primjenjuje u različitim disciplinama, uključujući kardiologiju, neurologiju i psihijatriju (Alzheimerova bolest, Parkinsonova bolest, Creutzfeldt-Jakobova bolest i shizofrenija), te već spomenutu epileptologiju. Različiti utjecajni čimbenici ograničavaju korištenje nelinearnih mjera za karakterizaciju dinamike EEG-a, ali kada se tumače s oprezom, pouzdano se mogu okarakterizirati različita stanja normalne i patološke funkcije mozga [3].

Fokus i cilj ovog rada je razvoj računalnog radnog okvira koji ima mogućnost učitavanja i vizualizacije elektroencefalografskih signala iz normiranih zapisa u ovisnosti o različitim parametrima u svrhu jednostavnije obrade i usporedbe, izlučivanje većeg broja linearnih i nelinearnih značajki i njihova pohrana u datoteku, te provođenje razvrstavanja EEG zapisa pacijenata za konkretan poremećaj, a u tu svrhu izabrana je epilepsija. Pritom se koristi gotov besplatan alat za dubinsku analizu podataka, WEKA (*Waikato Environment for Knowledge Analysis*) [5]. Dubinska analiza podataka (engl. *data mining*) predstavlja

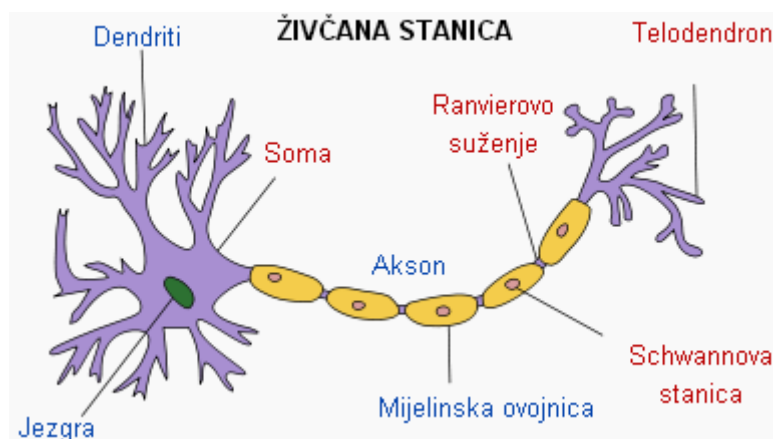
ekstrakciju nepoznatih, potencijalno korisnih informacija iz podataka. Uzorci i pravilnosti u podacima mogu se koristiti u budućim predviđanjima.

U radu se daje pregled osnovnih koncepata linearne i nelinearne EEG analize, a postupci korišteni u analizi epilepsije detaljnije su opisani. Posebna cjelina posvećena je opisu funkcionalnosti i načina korištenja razvijenog programskog produkta. Provedeni postupak testiranja za EEG zapise epilepsije detaljno je objašnjen, a rezultati su prikazani i komentirani.

2. Snimanje i pohrana elektroencefalograma

2.1. Ljudski mozak, prijenos akcijskog potencijala

Neuron ima važnu ulogu u prijenosu i obradi informacija u organizmu. Na temelju svoje strukture i funkcije može se podijeliti na tri glavna dijela: tijelo stanice (soma), kratki izdanci (dendriti) i dugačko živčano vlakno (akson). Neuron je prikazan na slici 1. Dendriti primaju podražaje iz drugih stanica i prenose ih u tijelo stanice, a akson potom prenosi signal iz tijela stanice do druge živčane ili mišićne stanice. Stanica je obavijena polupropusnim lipidnim dvoslojem, membranom debljine oko 7.5 - 10.0 nm. Važan dio stanične membrane su ionski kanalići (pore), kroz koje prolaze ioni natrija, kalija i klorida. Spoj između aksona i iduće stanice naziva se sinapsa. Dio sinapse na strani aksona naziva se presinapsa, a na strani susjedne stanice postsinapsa. Između njih postoji sinaptička pukotina debljine 10-50 nm. Prijenos impulsa putuje preko sinapse kao posljedica oslobađanja kemijskog transmittera u presinapsi, čime se, u slučaju dovoljno jakog podražaja, aktivira sljedeća stanica [6].



Slika 1. Građa živčane stanice ili neurona [<http://hr.wikipedia.org/wiki/Neuron>]

Napon membrane stanice definira se kao razlika potencijala unutarnje i vanjske površine membrane. Pritom se razlikuju potencijal mirovanja i promjene potencijala uslijed aktivnosti. U stanju mirovanja razlika potencijala nastaje zbog razlike koncentracije iona unutar i izvan stanice, pri čemu je koncentracija natrijevih iona (Na^+) oko 10 puta veća na vanjskoj strani membrane, nego na unutarnjoj strani, gdje je koncentracija kalijevih iona (K^+) oko 30 puta veća nego na vanjskoj strani membrane. Promjene potencijala mogu nastati kao

posljedica intrinzične aktivnosti stanice koja se javlja bez vanjske pobude, kao posljedica vanjske pobude, ili kao posljedica akcijskog potencijala, koji se javlja ako podražaj uzrokovan vanjskim djelovanjem prijeđe prag podražljivosti.

Stimuliranje živčane stanice dovodi do promjene transmembranskog napona. Stimulacija može biti ekscitacijska (depolarizacija, karakterizirana porastom potencijala unutar stanice, čime dolazi do pada normalnog negativnog napona mirovanja) ili inhibicijska (hiperpolarizacija, karakterizirana smanjenjem potencijala unutar stanice, čime dolazi do porasta potencijala u mirovanju). Nakon stimulacije membranski se napon vraća u originalnu vrijednost mirovanja. Ako je ekscitacijski podražaj membrane dovoljno jak, transmembranski potencijal doseže prag i membrana stvara karakteristični električni impuls, koji se naziva akcijskim impulsom ako aktivirana membrana slijedi zakon „sve ili ništa“, što znači da će njegova veličina uvijek biti jednaka, te je značajno samo da će se pojaviti ako je podražaj veći od praga, ili neće uopće. Inhibicijski podražaj povećava broj istodobnih ekscitacijskih podražaja potrebnih za dostizanje praga podražljivosti [7].

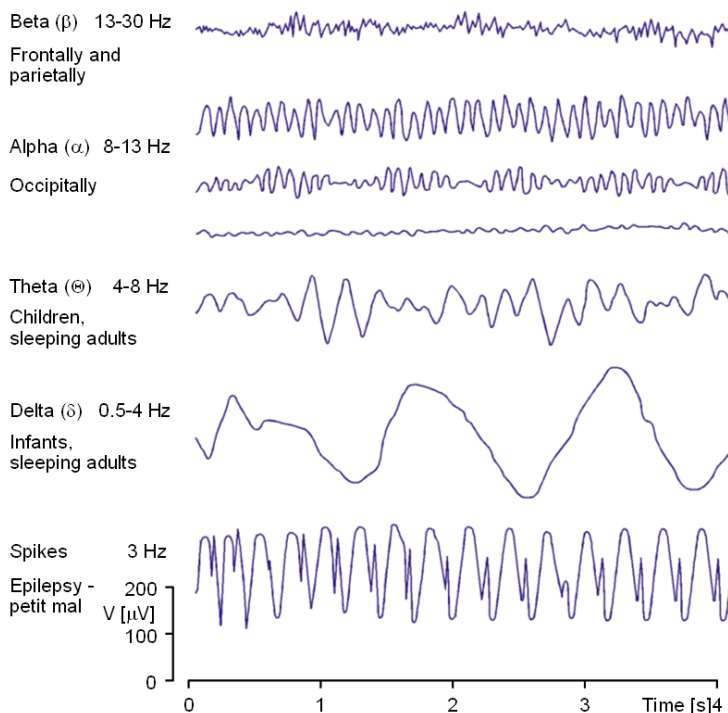
Pojava akcijskog potencijala i daljnji tijek događaja opisani su u nastavku. Kada je membrana stimulirana tako da transmembranski potencijal poraste otprilike 20 mV i dosegne prag, odnosno kada se membranski napon promijeni od -70 mV na oko -50 mV, mijenja se propusnost za ione natrija i kalija. Propusnost za ione natrija povećava se na početku vrlo brzo, što omogućava prolazak iona natrija izvana prema unutra kroz membranu, čime unutrašnjost postaje pozitivnija. Unutarnja strana membrane tada doseže potencijal od oko +20 mV. Nakon tog sporije rastuća propusnost za ione kalija omogućava ionima kalija prolazak iznutra prema van, a propusnost za Na^+ ione pada, vraćajući potencijal s unutarnje strane membrane na vrijednost mirovanja. Potom dolazi do smanjenja propusnosti za K^+ ione. Tijekom mirovanja nakon aktivacije Na-K pumpa (aktivni proces koji vraća K^+ izvan stanice u stanicu, a Na^+ izbacuje iz stanice) vraća ionske koncentracije s unutarnje i vanjske strane membrane na njihove originalne vrijednosti. Najveći porast transmembranskog napona tijekom aktivacije je oko 100 mV, a trajanje živčanog impulsa je oko 1 ms [6].

2.2. Elektroencefalografija

Prvo uspješno snimanje električnog polja ljudskog mozga proveo je 1924. njemački psihijatar i neurolog Hans Berger, a zapis koji je nastao snimanjem nazvao je elektroencefalogram. Od 1929. do 1938. objavio je 20 znanstvenih radova o EEG-u pod nazivom „Über das Elektrenkephalogramm des Menschen“ [8].

Zapis spontane aktivnosti mozga, koja se kontinuirano odvija u živućem pojedincu, mjerena na površini glave ima amplitudu oko 100 μV , a na površini mozga oko 1-2 mV. Osim spontane aktivnosti, moguće je mjeriti evocirane potencijale, koji predstavljaju komponente EEG-a koje nastaju kao reakcija na podražaj (električni, vizualni, auditorni i slično). Evocirani su potencijali obično ispod razine šuma i teško se razlučuju, zbog čega je potrebno koristiti metode obrade (metoda usrednjavanja) kako bi se poboljšao omjer signal-šum. Korištenjem mikroelektroda može se ispitati i ponašanje jednog neurona [7].

U EEG-u se mogu uočiti signali karakteristični po valnom obliku, veličini amplitude i frekvencijskom pojasu, a to su alpha (α), beta (β), delta (δ), i theta (Θ) valovi (Slika 2).



Slika 2. Primjeri EEG valova, redom α , β , δ , Θ , te šiljci povezani s epilepsijom [7]

Alpha valovi imaju frekvenciju između 8 i 13 Hz i dominiraju u zatiljnom režnju osobe u budnom opuštenom stanju. Amplituda im je do 50 μV . Nestaju za vrijeme spavanja, kao i prilikom koncentriranog rješavanja problema.

Beta valovi imaju frekvenciju između 13 i 30 Hz i mjerljivi su u tjemenom i čeonom režnju. Amplituda im iznosi do 20 μV . Javljaju se za vrijeme duševne aktivnosti, a oni najviših frekvencija za vrijeme duševne napetosti.

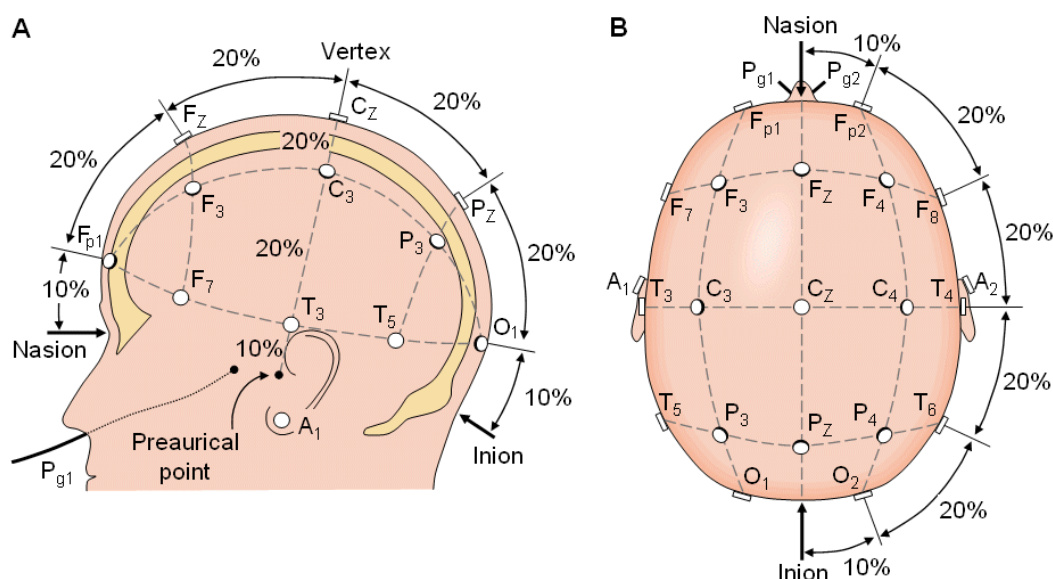
Delta valovi imaju raspon frekvencije od 0.5 do 4 Hz, a mogu se izmjeriti kod male djece i odraslih osoba u stanju dubokog sna, te pri težim bolestima mozga. Amplituda im je iznosa 60 μV - 100 μV . Pojavljuju se u kori velikog mozga neovisno o aktivnosti u nižim slojevima mozga.

Theta valovi imaju raspon 4-8 Hz i mjere se kod djece i odraslih osoba za vrijeme emocijskih stresova, posebno razočaranja i frustracija. Amplituda im iznosi do 70 μV , a pretežno se javljaju u tjemenom i sljepoočnom režnju. Mogu biti i patološki, te se pojavljuju kod mnogih poremećaja mozga [6].

2.3. Mjerenje EEG-a (međunarodni sustav 10-20)

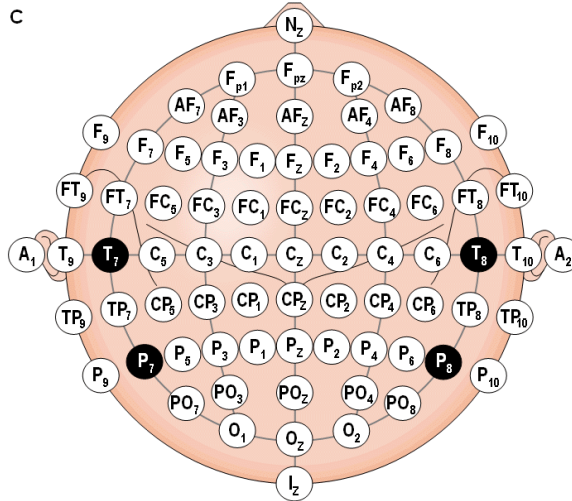
Da bi se omogućila usporedba dobivenih rezultata snimanja između različitih laboratorija potrebno je dobro definirati pozicije prema kojima su elektrode razmještene na glavi. Postoji više sustava postavljanja elektroda, a za snimanje spontane aktivnosti mozga raspored postavljanja i nomenklatura elektroda najčešće se odabiru prema normi 10-20. Sustav 10-20 sastoji se od 21 elektrode razmještene po unaprijed definiranim pozicijama na površini glave, pri čemu se „10“ i „20“ odnose na razmak elektroda, koji je ili 10 % ili 20 % ukupne udaljenosti lubanje, uzimajući na taj način u obzir varijacije u veličini glave (Slika 3). Pozicije se određuju tako da se mjeri od referentnih točaka, a to su tzv. *nasion*, udubljeno područje na vrhu nosa u razini s očima, i tzv. *inion*, koštana izbočina na najnižoj točki lubanje na središnjoj liniji sa stražnje strane glave. Sustav koristi specifičnu nomenklaturu za označavanje pozicija pojedinih elektroda s obzirom na građu mozga. Mozak je podijeljen na dvije polovice, lijevu i desnu hemisferu. Svaka od njih podijeljena je na čeonu (engl. *frontal*), tjemeni (engl. *parietal*), sljepoočni (engl. *temporal*) i zatiljni (engl. *occipital*) režanj. Za svaku

elektrodu postoji slovo koje identificira režanj i broj koji određuje lokaciju na polutki. Slova F, T, C, P i O se odnose na čeonu, sljepoočnu, „središnju“ (engl. *central*), tjemeni i zatiljni režanj, respektivno. Pritom je potrebno naglasiti da središnji režanj zapravo ne postoji, već se slovo „C“ koristi samo u identifikacijske svrhe. Oznaka „z“ (engl. *zero*) odnosi se na elektrode smještene na središnjoj liniji. Parni brojevi (2, 4, 6, 8) odnose se na pozicije elektroda na desnoj hemisferi, dok se neparni brojevi (1, 3, 5, 7) odnose na one na lijevoj hemisferi mozga.



Slika 3. Razmještaj 21 elektrode prema sustavu 10-20 s lijeve strane (A) i odozgo (B) [7]

Kada se snima detaljniji EEG s više elektroda, koriste se dodatne elektrode upotrebom dodatne 10 % podjele, tako da se ispunjavaju središnja mjesta na pola puta između elektroda postojećeg sustava 10-20. Kompliciraniji sustav novih pozicija i nomenklatura ovih elektroda doveo je do tzv. *Modified Combinatorial Nomenclature* (MCN). MCN sustav koristi oznake 1, 3, 5, 7, 9 za lijevu hemisferu, koje predstavljaju redom 10 %, 20 %, 30 %, 40 %, 50 % inion-nasion udaljenosti. Dolazi do preimenovanja četiri elektrode sustava 10-20: T₃, T₄, T₅ i T₆ nazvane su redom T₇, T₈, P₇ i P₈. Dodane su nove oznake: AF - između F_p i F, FC - između F i C, FT - između F i T, CP - između C i P, TP - između T i P, PO - između P i O (Slika 4).



Slika 4. Pozicija i nomenklatura između dodanih 10 % elektroda [8]

EEG je moguće mjeriti monopolarno ili bipolarno. Kod monopolarnog snimanja potencijal svake elektrode uspoređuje se neutralnom elektrodom (za koju se pretpostavlja da je na području koje nije zahvaćeno snimanjem) ili s prosjekom svih elektroda. Kod bipolarnog snimanja mjeri se razlika potencijala između dvije aktivne elektrode. Tako primjerice „F_{p1}-F₃“ predstavlja razliku u potencijalu između elektrode F_{p1} i elektrode F₃ [7].

2.4. EDF format

EDF format (engl. *European Data Format*) uobičajeni je format za pohranu i razmjenu višekanalnih bioloških i fizikalnih signala, koji pritom mogu imati bilo koje, čak i različite, fizikalne veličine i frekvencije uzorkovanja. Razvilo ga je nekoliko europskih inženjera medicine koji su se prvi put sastali 1987. u Kopenhagenu, kada se održavao tzv. *Sleep congress*. Inženjeri su od 1989. do 1992. radili na projektu tzv. *Sleep-Wake* analize i željeli su primijeniti svoje algoritme za analizu spavanja na međusobne podatke i usporediti rezultate analiza. U ožujku 1990. u Leidenu su se dogovorili oko vrlo jednostavnog zajedničkog formata podataka, koji je postao poznat kao EDF. Opis EDF formata objavljen je 1992. u časopisu „*Electroencephalography and Clinical Neurophysiology*“ [9]. Specifikacija se od tad nije mijenjala i još uvijek ima verziju 0. Danas se široko koristi u različitim istraživačkim projektima. Jedna EDF datoteka sadrži jedan neprekinuto digitalizirani poligrafski zapis. Sastoji se od zaglavlja (engl. *header*) iza kojeg slijede zapisi podataka (engl. *data record*).

Zapis u zaglavlju, koje je promjenjive duljine, sadrži informacije o pacijentu i specificira tehničke karakteristike snimljenih signala. Prvih 256 bajtova zaglavlja specificira broj verzije formata, identifikaciju pacijenta i snimanja, informaciju o vremenu i datumu početka snimanja, broj i trajanje zapisa podataka, te broj signala (n_s) u svakom zapisu podataka. Zatim slijedi 256 bajtova za svaki signal, koji specificiraju relevantne podatke o pojedinom signalu, kao što su tip signala, podaci o amplitudi, podaci o broju uzoraka u zapisu podataka, iz čega se može izvesti frekvencija uzorkovanja, jer je poznato trajanje svakog zapisa podataka. Zaglavlje sadrži $256 + (n_s * 256)$ bajtova. Vrijednost svakog uzorka reprezentirana je kao dvobajtni integer u formatu dvojnog komplementa. Mnogo detaljnija specifikacija EDF-a objavljena je 2003. pod nazivom EDF+ [10]. Formati su međusobno kompatibilni, tako da postojeći EDF preglednici ipak mogu prikazati EDF+ signale. EDF+ format je puno fleksibilniji, jer podržava i zapise koji nisu kontinuirani, kao što su tekstualne bilješke, podražaji i događaji u UTF-8 formatu. Zahvaljujući tome, svi signali, bilješke i događaji zabilježeni u jednom snimanju mogu se čuvati zajedno u jednoj datoteci.

3. Izlučivanje značajki iz EEG-a

Termin „vremenski niz“ u nastavku predstavlja niz mjerenja jedne varijable tijekom vremena (jedan signal). Kada je kontekst istodobno mjerenje više varijabli, koristi se naziv „multivarijatni vremenski nizovi“. Analiza uzoraka varijacije vremenskih nizova (tzv. *analiza varijabilnosti*) može se definirati kao opsežna procjena o stupnju i karakteru obrazaca varijacija tijekom vremenskih intervala. Dvije glavne kategorije tehnika analize varijabilnosti su transformacije i postupci za izlučivanje značajki. Zajedničko im je rukovanje podacima kako bi se identificirala određena interesantna svojstva. Značajke signala predstavljaju određeni dio informacije koji se može izlučiti iz potencijalno dostupnih informacija u signalu primjenom nekog od postupaka za izlučivanje značajki. Iz EEG zapisa mogu se prepoznati i izdvojiti brojne značajke, a ako je potrebno (i moguće s obzirom na količinu i kvalitetu podataka) i pratiti njihov vremenski razvoj. Postupci za izlučivanje značajki nisu neovisni o transformacijama: u mnogim slučajevima potrebno je primijeniti transformacije na podatke prije izlučivanja značajki. Naime, transformacije su funkcije koje preslikavaju uzorke skupa podataka iz jednog skupa, odnosno prilagođavaju podatke kako bi se nakon toga mogle dobiti možda bitnije značajke nego iz izvornih podataka, inače skrivene ili nedostupne. Neki od primjera tehnika transformacija su STFT (engl. *short-time Fourier transform*), WVT (engl. *Wigner-Ville transform*) i WT (engl. *wavelet transform*), koje spadaju u vremensko-frekvencijske transformacije, te tzv. *recurrence plots* i tzv. *Poincaré plots* [11]. Nakon izlučivanja značajki često se provodi postupak klasifikacije. Cijeli postupak možemo zamisliti kao preslikavanje N -dimenzionalnog prostora EEG zapisa u F -dimenzionalni prostor uzoraka (gdje je $N \geq F$), nakon čega se u prostoru uzorka pomoću klasifikatora dobivaju rezultati (primjerice razlikuju faze sna ili određuje radi li se o epileptičkom napadaju).

3.1. Koncept dinamičkog sustava

Dinamički sustav je model opisan stanjem (vrijednosti svih varijabli koje opisuju sustav u određenom vremenskom trenutku) i dinamikom (promjena stanja sustava u vremenu). Stanje je opisano s m varijabli koje se mogu prikazati kao točka u m -dimenzionalnom prostoru, kojeg zovemo prostor stanja ili fazni prostor sustava. Dinamika sustava je skup

zakona ili jednadžbi koje opisuju kako se stanje sustava mijenja tijekom vremena. Dinamički razvoj sustava odgovara nizu uzastopnih stanja (točaka) u prostoru stanja, a linija koja povezuje te uzastopne točke u prostoru stanja naziva se putanja (trajektorija) sustava.

Dinamički sustavi mogu se podijeliti na više načina. Jedan od njih je već spomenuta podjela na linearne (sve jednadžbe koje opisuju sustav su linearne, linearna veza između uzroka i posljedica) i nelinearne sustave (nisu sve jednadžbe koje opisuju sustav linearne, mali uzrok može imati veliki utjecaj). Dinamički sustavi mogu se podijeliti i na konzervativne (sustav bez otpora ili gubitka energije tijekom vremena) i disipativne (postoji otpor ili gubitak energije). Osim toga, mogu se podijeliti na determinističke (jednadžbe gibanja ne sadrže nikakav šum u uvjetima) i stohastičke.

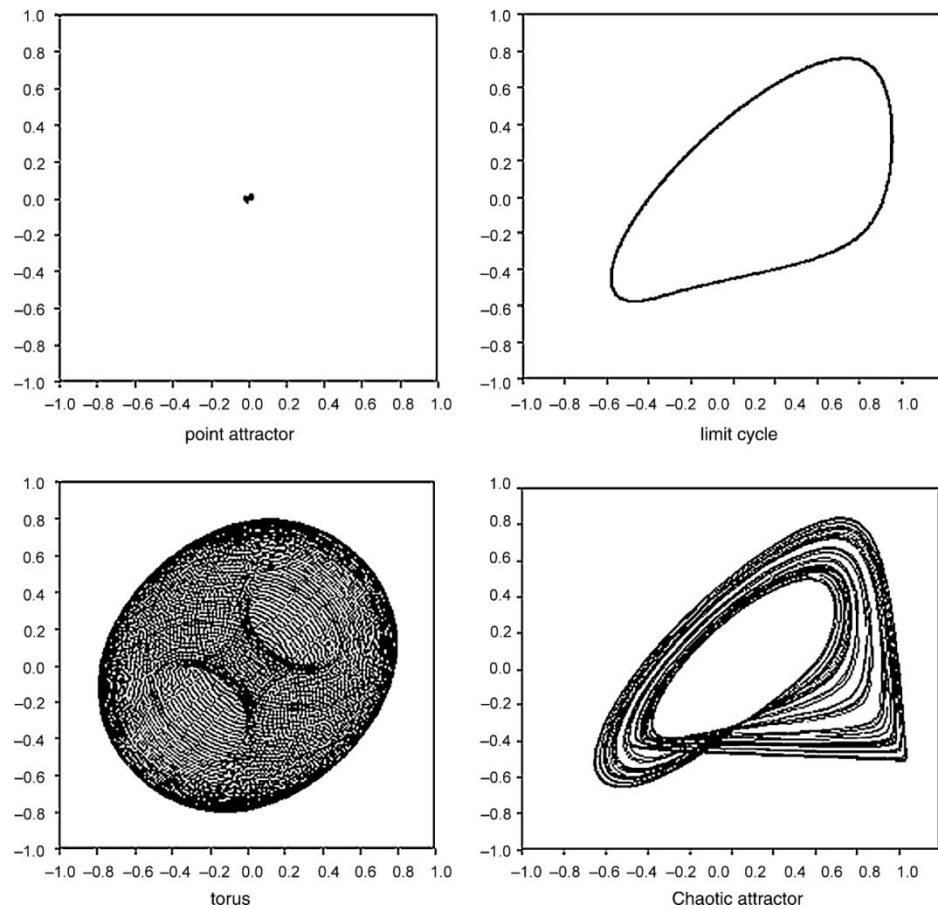
Realistični biološki sustavi (npr. neuronske mreže u mozgu) uglavnom su nelinearni disipativni sustavi, a nelinearna se analiza bavi problemom jesu li više deterministički ili stohastički [4].

3.1.1. Nelinearni dinamički sustavi

3.1.1.1. Atraktori

Osnovno svojstvo disipativnih determinističkih dinamičkih sustava je konvergiranje trajektorije u potprostor ukupnog prostora stanja ako se sustav dovoljno dugo promatra (nakon istitravanja početnih prijelaza). Taj potprostor je geometrijski objekt koji se zove atraktor sustava, jer „privlači“ trajektorije svih mogućih početnih uvjeta.

U linearnim disipativnim determinističkim sustavima može postojati samo jedan tip atraktora – jednostavna točka u prostoru stanja, odnosno točkasti atraktor (engl. *point attractor*), iz čega slijedi da će sustav konvergirati u stabilno stanje nakon kojeg se više neće događati daljnje promjene, osim ako se na sustav ne djeluje izvana. Kod nelinearnih determinističkih disipativnih sustava mogu se javiti uz točkaste atraktore još tri tipa atraktora: periodičan atraktor (engl. *limit cycle*), atraktor u obliku prstena kao rezultat kvaziperiodičnosti (engl. *torus attractor*) i neobičan (kaotičan) atraktor (engl. *strange (chaotic) attractor*), prikazani na slici 5.



Slika 5. Vrste atraktora [4]

Dinamika koja odgovara kaotičnom atraktoru je deterministički kaos. To je vrsta dinamike koja je s jedne strane deterministička, ali s druge je strane naizgled slučajna. Kaotična dinamika može se predvidjeti samo za kratke periode. Kaotični sustav, iako je njegova dinamika ograničena atraktorom, nikad ne ponavlja isto stanje. Taj paradoks je moguć zbog fraktalne strukture atraktora. Atraktori su vrlo važni objekti jer daju sliku dinamike sustava; što je atraktor kompleksniji, kompleksnija je odgovarajuća dinamika [4].

3.1.1.2. Svojstva atraktora

Za karakterizaciju svojstva atraktora koristi se više mjerenja, kao što su dimenzija atraktora, Lyapunovljevi eksponenti i entropija.

Dimenzija geometrijskog objekta mjeri je njegove prostorne opsežnosti. Dimenzija atraktora može se promatrati kao mjera stupnjeva slobode ili „složenosti“ dinamike. Točkasti

atraktor ima dimenziju 0, periodični atraktor dimenziju 1, torus ima cjelobrojnu dimenziju koja odgovara nametnutim periodičnim oscilacijama, a neobični atraktor ima fraktalnu dimenziju, koja nije cijeli broj. Dimenzije su statičke mjere atraktora, koje ne pružaju informaciju o evoluciji trajektorija tijekom vremena.

Lyapunovljevi eksponent i entropija smatraju se „dinamičkim“ mjerama složenosti atraktora. Lyapunovljevi eksponenti ukazuju na eksponencijalnu divergenciju (pozitivni eksponenti) ili konvergenciju (negativni eksponenti) obližnjih trajektorija prema atraktoru. Sustav ima onoliko Lyapunovljevih eksponenata koliko ima smjerova u prostoru stanja. Kontinuirani dinamički sustavi uvijek imaju bar jedan eksponent koji je točno nula. Koncept entropije blisko je povezan s Lyapunovljevim eksponentima. Kolmogorovljeva entropija je definirana kao mjera gubitka informacija tijekom vremena i procijenjena je sumom svih pozitivnih Lyapunovljevih eksponenata.

Kaotična dinamika može se definirati u terminima Lyapunovljevih eksponenata i entropije: karakteriziraju je postojanje bar jednog pozitivnog Lyapunovljevog eksponenta, odnosno, ekvivalentno, pozitivna entropija. Pozitivan Lyapunovljev eksponent ili entropija odražavaju tendenciju eksponencijalnog rasta malih poremećaja. Zato se govori o „osjetljivoj ovisnosti o početnim uvjetima“ u kaotičnim sustavima i ograničava se njihova mogućnost predviđanja [4].

3.1.1.3. Kontrolni parametri, višestruka stabilnost, bifurkacije

Kontrolni parametri su svojstva sustava koja mogu utjecati na dinamiku sustava i koja su ili nepromijenjena ili se pretpostavljaju konstantnima tijekom vremena dok je sustav promatran. Za fiksiran skup kontrolnih parametara dinamički sustav može imati više od jednog atraktora. Taj fenomen naziva se višestruka stabilnost. Svaki atraktor zauzima svoje područje u prostoru stanja sustava. Ako početno stanje sustava padne unutar područja privlačnosti određenog atraktora, dinamika sustava evoluirat će prema tom atraktoru i ostati tamo. Tako će u sustavu s višestrukom stabilnosti područja privlačnosti odrediti na kojem atraktoru će završiti sustav. Vanjski utjecaji mogu „gurnuti“ sustav izvan područja privlačnosti jednog atraktora i pomaknuti ga u područje drugog. Kod kritičnih vrijednosti

kontrolnih parametara moguća je pojava iznenadnih i dramatičnih prijelaza, koje nazivamo bifurkacije, kod kojih stari atraktori mogu nestati, a novi se pojaviti (primjer epileptičkih napadaja koji nastaju odjednom usred normalne aktivnosti mozga) [4].

3.2. Nelinearna analiza vremenskih nizova

Pomoću nelinearne analize vremenskih nizova možemo promatrajući sustav s nepoznatim svojstvima doći do boljeg razumijevanja dinamike osnovnog sustava. Jedan od pristupa unutar nelinearne analize je procedura koja se sastoji od tri posebna koraka [4]:

1. rekonstrukcija dinamike sustava u prostoru stanja
2. karakterizacija rekonstruiranog atraktora
3. provjera valjanosti procedure testiranjem sa zamjenskim podacima

3.1.1. Rekonstrukcija dinamike iz promatranja – ugrađivanje (engl. *embedding*)

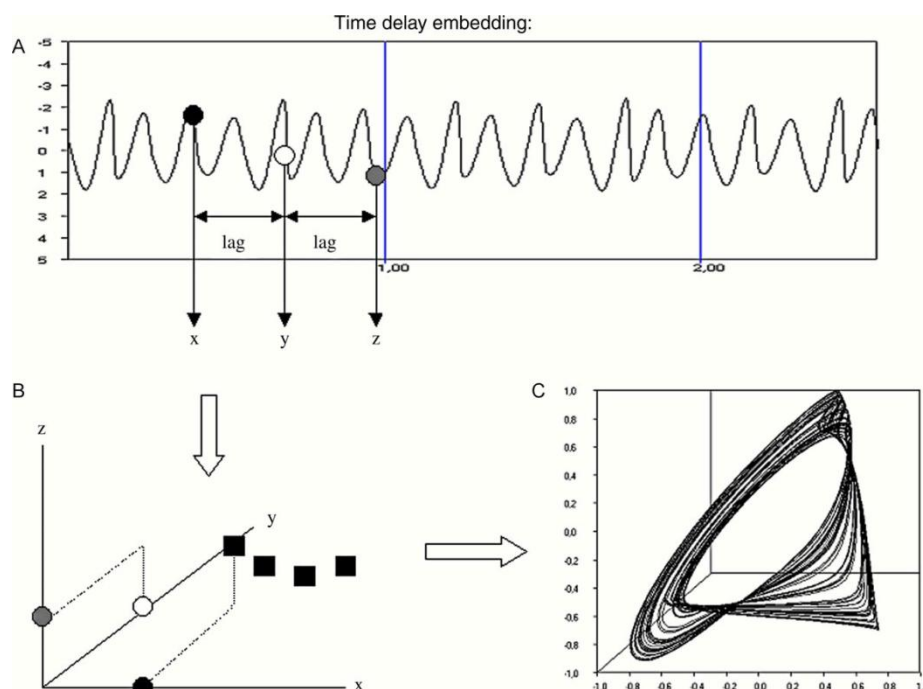
Rekonstrukcija atraktora u prostoru stanja osnovnog sustava (iz jednog ili više vremenskih nizova) predstavlja prvi i najznačajniji korak u nelinearnoj analizi. Kod ugrađivanja jedan ili nekoliko istovremenih vremenskih nizova pretvaraju se u niz ili slijed vektora u m -dimenzionalnom prostoru ugrađivanja. Ako sustav iz kojeg su uzeta mjerenja ima atraktor i ako je prostor ugrađivanja dovoljno velik (više od dvostruko veći od dimenzije atraktora sustava), niz rekonstruiranih vektora čini „ekvivalentni atraktor“. Takens je dokazao da ekvivalentni atraktor ima ista dinamička svojstva (dimenziju, Lyapunovljeve eksponente, entropiju) kao stvarni atraktor („Takensov teorem ugrađivanja“) [12]. Iz tog slijedi da možemo dobiti vrijedne informacije o dinamici sustava, čak i ako nemamo izravan pristup svim varijablama sustava. Postoje dva različita postupka ugrađivanja:

1. vremenski odgođeno ugrađivanje
2. prostorno ugrađivanje

3.1.1.1. Vremenski odgođeno ugrađivanje (engl. *time-delay embedding*)

Kod vremenski odgođenog ugrađivanja započinje se s jednim vremenskim nizom promatranja, iz kojeg se rekonstruiraju m -dimenzionalni vektori uzimanjem m uzastopnih vrijednosti vremenskog niza kao vrijednosti za m koordinata vektora. Ponavljanjem postupka

za sljedećih m vrijednosti vremenskog niza, dobiva se niz vektora u prostoru stanja sustava. Veza između susjednih vektora definira trajektoriju sustava. U praksi se ne koriste vrijednosti uzastopnih koraka digitalizacije vremenskog niza, već se uzimaju vrijednosti odvojene malim odmakom l . Tako je ovo ugrađivanje karakterizirano s dva parametra: vremenskim odmakom l i dimenzijom ugrađivanja m . Prikladan odabir tih parametara je važan, ali težak korak u nelinearnoj analizi. Pragmatičan pristup je odabrati l jednak vremenskom intervalu nakon kojeg autokorelacijska funkcija (ili zajednička informacija) vremenskog niza padne na $1/e$ početne vrijednosti, i ponoviti analizu (primjerice računanje korelacijske dimenzije) za rastuće vrijednosti m dok se rezultati ne prestanu mijenjati; pretpostavlja se da je to točka gdje je $m > 2d$ (gdje je d stvarna dimenzija atraktora). Postoje i sofisticiraniji postupci za odabir odmaka [13], kao i odabir dimenzije ugrađivanja [14]. Bez obzira na odabrani postupak, važno je znati da su l i m međuzavisni. Produkt l i m , nazvan prozorom ugrađivanja, dužina je segmenta vremenskog niza korištenog za rekonstrukciju jednog vektora u prostoru stanja. Takens predlaže da se l odabire tako da uhvati najmanji detalj od interesa u vremenskoj seriji, a m tako da prozor ugrađivanja uhvati najveći fenomen od interesa. Primjer je prikazan na slici 6.



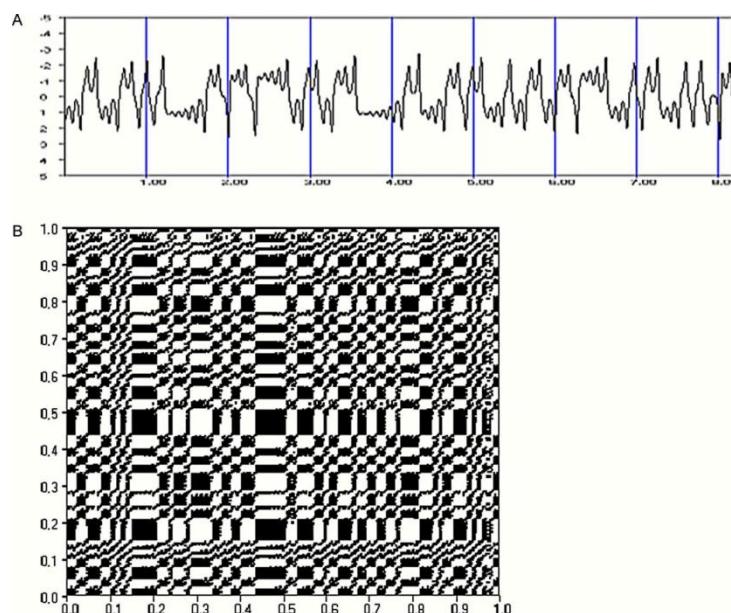
Slika 6. Vremenski odgođeno ugrađivanje. Pod A je prikazan mali segment vremenskog niza. Da bi se vremenski niz ugradio u trodimenzionalni prostor, potrebno je uzeti tri koordinate (amplitudne vrijednosti) u tri uzastopna trenutka, x , y i z , odvojene razmakom lag . Prostor između x i z naziva se prozor ugrađivanja. Dobiva se koordinata točke u prostoru stanja prikazana kvadratom (pod B), a ponavljanjem ovog postupka za vrijednosti u sljedećim trenucima dobiva se iduća točka u prostoru stanja. Ekvivalentni atraktor, prikazan pod C, može se rekonstruirati ponavljanjem postupka za cijeli vremenski niz [4].

3.1.1.2. Prostorno ugrađivanje (engl. spatial embedding)

Ukoliko je dostupno m vremenskih serija neovisnih mjerenja umjesto samo jednog, moguće je koristiti prostorno ugrađivanje za rekonstrukciju atraktora sustava. U ovom slučaju m koordinata vektora se uzima kao vrijednosti m vremenskih nizova u određenom trenutku, a ponavljanjem ovog postupka za uzastopne vremenske trenutke dobije se niz vektora. Dimenzija ugrađivanja m jednaka je broju kanala korištenih za rekonstrukciju vektora. Prostorni ekvivalent vremenskog odmaka l je udaljenost između elektroda. Prednost prostornog ugrađivanja je postizanje značajne redukcije podataka, budući da je dinamika cijelog sustava prikazana u jednom prostoru stanja. Alternativa bi bila napraviti odvojeno vremenski odgođeno ugrađivanje za svaki od m vremenskih nizova. Nedostatak ovog pristupa je to što je prostorni odmak (udaljenost između EEG elektroda ili MEG senzora), uglavnom zadan i ne može se odabrati na optimalan način. Nije sigurno je li prostorno ugrađivanje valjani postupak ugrađivanja, iako ga neki snažno zagovaraju [4].

3.1.2. Karakterizacija rekonstruiranog atraktora

Nakon rekonstrukcije atraktora idući korak je karakterizacija. Najjednostavniji način je vizualizacija pomoću transformacija kao što su fazni portret (engl. *phase portrait*) ili Poincaréova sekcija (engl. *Poincaré section*). Neki Poincaréovu sekciju (transformaciju) poistovjećuju s omjerom SD1/SD2 u faznom prostoru što predstavlja zapravo samo jednu od mogućih značajki. Mnogo složeniji, ali vrlo informativan način za prikaz rekonstruiranih segmenata putanje je tzv. *recurrence plot* (slika 7). Na recurrence plotu se definira obično više značajki (četiri ili pet). Nakon ugrađivanja i moguće vizualizacije rekonstruiranog atraktora provode se postupci kojima se pokušava napraviti karakterizacija na kvantitativan način, kao što su korelacijska dimenzija, Lyapunovljevi eksponenti, entropija itd.



Slika 7. Primjer recurrence plot. Na slici B prikazan je recurrence plot, 2D graf čije obje osi odgovaraju vremenu, iscrtan na temelju vremenskog niza sa slike A. Svaka točka u grafu odgovara kombinaciji dva vremena (vrijednosti x i y koordinata). Kad su vektori prostora stanja koji odgovaraju tim vremenskim točkama bliži nego neka mala granična udaljenost, točka se prikazuje crnom u grafu, inače se ne crta [4].

3.1.3. Provjera valjanosti testiranjem sa zamjenskim podacima

Interpretacija nelinearnih mjera ponekad može predstavljati problem, jer filtrirani šumoviti vremenski nizovi mogu dovesti do pogrešnog dojma niskodimenzionalne dinamike i kaosa, a testiranje zamjenskim podacima (engl. *surrogate data testing*) predstavlja jedan od najvažnijih alata zaštite. Osnovni pristup je jednostavan: nelinearne mjere, kao što su

primjerice entropija ili korelacijska dimenzija, računaju se iz vremenskog niza i iz zamjenskih podataka. Podaci zamjenskog vremenskog niza konstruiraju se tako da imaju ista linearna svojstva (spektar snage/autokorelacijska funkcija) kao originalni vremenski niz, ali bez ikakve druge (nelinearne) strukture. Ukoliko je rezultat nelinearne analize jasno različit za originalne i zamjenske podatke, može se zaključiti da originalni podaci sadrže zanimljivu nelinearnu strukturu. Elegantan način konstruiranja zamjenskih podataka jednakog spektra snage je provođenje Fourierove transformacije, zatim ispremiješanje faza i nakon tog provođenje obrnute Fourierove transformacije. Ipak, čak i uz testiranje pomoću zamjenskih podataka, postoji mogućnost pogrešnih rezultata [4].

3.2. Pregled implementiranih EEG značajki

Postupci za izlučivanje značajki mogu se klasificirati u različite grupe tehnika koje pružaju slične načine za izračunavanje i sličan sadržaj informacija unutar grupe. Grupe su tradicionalno označavane kao domene varijabilnosti. Iako je prošlo neko vrijeme, klasifikacija tehnika varijabilnosti nije značajno evoluirala od važnog rada Radne skupine (engl. *Task force*) za varijabilnost srčanog ritma [15].

Task Force je 1996. između ostalog specificirao standardne postupke mjerenja, koje je podijelio u 3 domene: vremensku, unutar koje se nalaze statistički i geometrijski postupci, frekvencijsku i općenitiju nelinearnu domenu. Međutim, odabrani postupci iz vremenske i frekvencijske domene navedeni u radu Task Forcea odnose se isključivo na varijabilnost srčanog ritma. U nelinearnu domenu svrstane su Poincaréove sekcije, niskodimenzionalni dijagrami atraktora i trajektorije atraktora, te korelacijska dimenzija D_2 , Lyapunovljevi eksponenti i Kolmogorovljeva entropija, koji se također primjenjuju i u analizi EEG-a.

Seely je 2005. ponudio specifičniju klasifikaciju tehnika varijabilnosti, koja razlikuje nekoliko domena varijabilnosti, kao što su vremenska domena, frekvencijska domena, domena entropije i domena invarijantna s obzirom na vremensko-prostornu skalu [16]. Analiza u vremenskoj domeni predstavlja najjednostavniji način vrednovanja varijabilnosti, identificirajući mjere varijacije tijekom vremena, kao što je standardna devijacija. Analiza u frekvencijskoj domeni prikazuje doprinose svakog sinusnog vala kao funkciju njegove

frekvencije, a primjer je spektralna gustoća snage. U domeni invarijantnoj s obzirom na vremensko-prostornu skalu spada postupak kolebanja s uklonjenim trendom (engl. *detrended fluctuation analysis, DFA*). Primjeri postupaka iz domene entropije su približna entropija (ApEn) i entropija uzorka (SampEn).

Iako se ovo razvrstavanje pokazalo korisnim, ono ne obuhvaća sve tehnike analize varijabilnosti dostupne danas. Bravi i koautori 2011. predstavljaju malo širu klasifikaciju postupaka za izlučivanje značajki, kao što je podjela na pet različitih domena varijabilnosti: statističku, za koju povezane značajke opisuju statistička svojstva distribucije i koja pretpostavlja da su podaci iz stohastičkog procesa; geometrijsku, koja opisuje ona svojstva koja se odnose na oblik skupa podataka u određenom prostoru; energetske, koja opisuje značajke koje se odnose na energiju ili snagu vremenskog niza; informacijsku, koja opisuje stupanj nepravilnosti, tj. složenosti svojstvenu redoslijedu elemenata u vremenskoj seriji; invarijantu, koja opisuje svojstva sustava koji pokazuju fraktalnost ili druge osobine koje se ne mijenjaju tijekom vremena ili prostora [11]. Ovakvom klasifikacijom vremenska je domena podijeljena na statističku i geometrijsku, kao što je već Task Force predložio. Frekvencijska domena je uključena sa značajkama poput energijskih operatora (engl. *energy operators*) u općenitiju energetske domenu, a entropijska i domena neovisna o skali postale su dio općenitijih, redom, informacijske i invarijantne domene. U statističku domenu mogu se svrstati značajke simboličke dinamike (engl. *symbolic dynamics features*) i standardne statističke značajke. U geometrijsku domenu spadaju značajke Poincaré plotova i značajke recurrence plotova. U energetske domenu spadaju energijski operatori, te frekvencijske i vremensko-frekvencijske značajke (engl. *time-frequency features*). U informacijskoj domeni nalaze se značajke kao što su ApEn i SampEn. U invarijantnu domenu spadaju značajke korelacijska dimenzija, DFA i najveći Lyapunovljev eksponent.

Razvijeni radni okvir sadrži veliki broj značajki, koje su navedene u tablici I, a one korištene u radu za analizu epilepsije ukratko su opisane u cjelini u nastavku.

TABLICA I. IMPLEMENTIRANI POSTUPCI I ZNAČAJKE

Značajke/parametri	Kategorija	Paket
Srednja vrijednost, standardna devijacija, srednja vrijednost apsolutnih vrijednosti prvih razlika, srednja vrijednost apsolutnih vrijednosti prvih razlika (normaliziranih signala), srednja vrijednost apsolutnih vrijednosti drugih razlika, srednja vrijednost apsolutnih vrijednosti drugih razlika (normaliziranih signala), autokorelacijski koeficijent, Fanov faktor [17, 18]	Linearne, vremenska domena, statističke	features.linear.timeDomain
AlphaPSD, BetaPSD, GammaPSD, DeltaPSD, ThetaPSD (5 značajki)/frekvencija uzorkovanja, FFT PSD (prozor - Hann, Hamming, ništa) ili Burgova PSD procjena (AR red modela) [17]	Linearne, frekvencijska domena	features.linear.frequency
Haar wavelet standardna devijacija/skala, Hilbert-Huangova transformacija [18, 19]	Vremensko-frekvencijska	features.timeFrequency
Najveći Lyapunovljev eksponent/(dimenzija, lag, r), korelacijska dimenzija/(dimenzija, lag, preciznost), greška nelinearnog predviđanja/(dimenzija, lag, T), indeks prostorne potpunosti (SFI)/(dimenzija, lag, preciznost), omjer standardnih devijacija (SD1/SD2), značajke recurrence plot - prosječan broj susjeda, recurrence rate, Lmean, DET, Shannonova entropija recurrencea, LAM (6 značajki)/(dimenzija, lag, r), CTM za točke faznog prostora/(dimenzija, lag, r) [20, 21, 22, 23, 24, 25, 26]	Nelinearne, fazni prostor	features.nonlinear.phaseSpace
Higuchijeva fraktalna dimenzija/ k_{max} , Hurstov eksponent, DFA – AlphaL, AlphaS (2 značajke)/min duljina segmenta, granica za AlphaL, zastavica za računanje AlphaL [27, 18, 28]	Nelinearne, fraktalne	features.nonlinear.fractal
ApEn/(m faktor, r), max ApEn/ r , SampEn/(m faktor, r), max SampEn/ r , Rényi/red [29, 30, 31]	Nelinearne, entropija	features.nonlinear.entropy
Složenost Lempel-Ziv, Allanov faktor/prozor promatranja, CTM/ r [32, 18, 34]	Nelinearne, ostale	features.nonlinear.other
Zajednička dimenzija/(dimenzija 1, dimenzija 2, lag 1, lag 2, preciznost), značajke cross recurrence plot - CRP recurrence rate, CRP Lmean, CRP DET, CRP Shannonova entropija recurrencea, CRP LAM (5 značajki)/(dimenzija, lag, r), vjerojatnost sinkronizacije/(indeksi signala, dimenzija, lag, recurrence number, ρ_{ref} , ϵ_{max}) [35, 25, 36]	Nelinearne, multivarijatne	features.nonlinear.multiSeries

3.3. Opis značajki korištenih u analizi epilepsije

3.3.1. Srednja vrijednost sirovih signala (engl. *mean*)

Srednja vrijednost signala dana je sljedećim izrazom:

$$\mu_x = \frac{1}{N} \sum_{n=1}^N X_n \quad (1)$$

X_n je amplituda n -tog uzorka vremenskog niza, a N ukupan broj uzoraka u promatranom segmentu [17].

3.3.2. Hurstov eksponent (engl. *Hurst exponent*)

Koristi se za kvantificiranje korelacije točaka u vremenskom nizu [18]. Pritom se računa tzv. *rescaled range* (R/S) za vremenski niz $X = X_1, X_2, \dots, X_n$ na sljedeći način:

1. Računa se srednja vrijednost zadana izrazom (1)
2. Oduzima se srednja vrijednost od svakog uzorka:

$$Y_t = X_t - \mu_x \quad (2)$$

za $t = 1, 2, \dots, n$.

3. Računaju se kumulativna odstupanja:

$$Z_t = \sum_{i=1}^t Y_i$$

za $t = 1, 2, \dots, n$.

4. Računa se $R(t)$:

$$R_t = \max(Z_1, Z_2, \dots, Z_t) - \min(Z_1, Z_2, \dots, Z_t) \quad (3)$$

za $t = 1, 2, \dots, n$.

5. Računa se standardna devijacija niza S :

$$S_t = \sqrt{\frac{1}{t} \sum_{i=1}^t (X_i - \mu_x)^2} \quad (4)$$

za $t = 1, 2, \dots, n$.

6. R/S se računa kao:

$$\left(\frac{R}{S}\right)_t = \frac{R_t}{S_t} \quad (5)$$

za $t = 1, 2, \dots, n$.

7. Hurstov eksponent dobiva se računanjem nagiba linije dobivene kao $\log(R/S)$ naspram $\log(T)$

3.3.3. Korelacijska dimenzija (engl. *correlation dimension*, D_2)

Temelji se na integralu povezanosti $C(\varepsilon)$, koji predstavlja vjerojatnost da će bilo koje dvije slučajno odabrane točke atraktora biti bliže nego zadana udaljenost ε [21] i računa se kao:

$$C(\varepsilon) = \frac{2}{(M-W)(M-W-1)} \sum_{i=1}^M \sum_{j=i+W}^M \theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \quad (6)$$

gdje $\|\dots\|$ označava neku normu u m dimenzija, a θ je Heavisideova step funkcija ($\theta(a) = 0$ za $a \leq 0$ i $\theta(a) = 1$ za $a > 0$). $M = N - (m - 1) \tau$, gdje je τ vremenski odmak (vremenski odgođeno ugrađivanje), a m je dimenzija ugrađivanja. Za dovoljno veliku dimenziju ugrađivanja m , nagib linearno skaliranog područja $\ln C(\varepsilon) / \ln \varepsilon$ je procjena korelacijske dimenzije D_2 , što je ključna točka algoritma. Da bi se odredio dovoljno veliki m , postupak se ponavlja za rastuće vrijednosti m dok vrijednost dimenzije povezanosti ne prestane rasti, što možemo zapisati kao:

$$D_2 = \lim_{N \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} d(\varepsilon), \quad (7)$$

gdje je

$$d(\varepsilon) = \frac{d \ln C(\varepsilon)}{d \ln \varepsilon} \quad (8)$$

3.3.4. Najveći Lyapunovljev eksponent (engl. *largest Lyapunov exponent*)

Najveći Lyapunovljev eksponent L_{max} može se procijeniti iz:

$$d_j(i) \approx C_j e^{L_{max} i \Delta t}, \quad (9)$$

gdje $d_j(i)$ označava prosječno odstupanje između dva segmenta trajektorije u vremenu t_i . C_j sa $j = 1, \dots, M$ je konstanta koja se dobije inicijalnim odvajanjem referentnog vektora u prostoru stanja i njegovog najbližeg susjeda. Logaritmiranjem obje strane jednadžbe dobiva se

$$\ln d_j(i) \approx \ln C_j + L_{max} * i * \Delta t \quad (10)$$

Najveći se Lyapunovljev eksponent onda računa kao nagib pravca dobivenog prema kriteriju najmanjih kvadrata krivulje:

$$y(i) = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle \quad (11)$$

po periodima promatranja razvoja trajektorije delta t , gdje $\langle \dots \rangle$ označava prosjek po svim trajektorijama j [20].

3.3.5. Entropija uzorka (engl. *sample entropy, SampEn*)

Definira se $B_i^m(r)$ kao $(N-m-1)^{-1}$ puta broj vektora X_j^m unutar r od X_i^m , gdje $j = 1, \dots, N-m$ i $j \neq i$ da bi se isključile usporedbe sa samim sobom i zatim se definira

$$B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r) \quad (12)$$

Slično se definira $A_i^m(r)$ kao $(N-m-1)^{-1}$ puta broj vektora X_j^{m+1} unutar r od X_i^{m+1} , gdje $j = 1, \dots, N-m$ i $j \neq i$ i postavi se

$$A^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} A_i^m(r) \quad (13)$$

Parametar $\text{SampEn}(m, r)$ se onda definira kao $\lim_{N \rightarrow \infty} \left\{ -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \right\}$, što se može procijeniti po statistici:

$$\text{SampEn}(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (14)$$

Postavi se $B = \left\{ \left[\frac{(N-m-1)(N-m)}{2} \right] \right\} B^m(r)$ i $A = \left\{ \left[\frac{(N-m-1)(N-m)}{2} \right] \right\} A^m(r)$, tako da je B ukupan broj poklapanja za početnu dimenziju m i A ukupan broj poklapanja za dimenziju $m+1$, i onda se SampEn može izraziti kao $-\ln \left(\frac{A}{B} \right)$ [30].

3.3.6. Mjera središnje težnje (engl. *central tendency measure, CTM*)

Pomoću CTM-a se, umjesto definiranja vremenskog niza kao kaotičnog ili nekaotičnog, procjenjuje stupanj varijabilnosti ili kaosa [34]. Ako je x_i amplituda n -tog uzorka vremenskog niza, a N ukupan broj uzoraka u promatranom segmentu, CTM se može izračunati kao:

$$CTM = \frac{\sum_{i=1}^{N-2} \delta(d_i)}{N-2}, \quad (15)$$

gdje je

$$\delta(d_i) = \begin{cases} 1, & \text{ako } \left[(x(i+2) - x(i+1))^2 + (x(i+1) - x(i))^2 \right]^{1/2} < \rho \\ 0, & \text{inače} \end{cases} \quad (16)$$

Polumjer ρ se odabire ovisno o svojstvu podataka. Niska vrijednost CTM-a ukazuje na veliku količinu rasipanja i visoku vrijednost koncentracije blizu središta.

3.3.7. Složenost Lempel-Ziv (engl. *Lempel-Ziv (LZ) complexity*)

Analiza LZ složenosti temelji se na transformiranju signala u konačan slijed simbola [32]. Srednja vrijednost se procjenjuje kao prag T_d . Usporedbom s pragom, podaci signala pretvaraju se u slijed nula i jedinica $P = s(1), s(2), \dots, s(n)$, gdje je $s(i)$ definiran kao:

$$s(i) = \begin{cases} 0 & \text{if } x(i) < T_d \\ 1 & \text{if } x(i) \geq T_d \end{cases} \quad (17)$$

Mjera složenosti može se procijeniti korištenjem sljedećeg algoritma:

1. Neka S i Q označavaju dva podniza P , a SQ neka je ulančavanje S i Q , dok je slijed $SQ\pi$ izveden iz SQ nakon brisanja posljednjeg znaka niza (π označava operaciju brisanja zadnjeg znaka u nizu). Neka $v(SQ\pi)$ označava vokabular svih različitih podnizova $SQ\pi$. Na početku, $c(n) = 1$, $S = s(1)$, $Q = s(2)$, dakle, $SQ\pi = s(1)$.
2. Općenito, ako $S = s(1), s(2), \dots, s(r)$, $Q = s(r+1)$, onda $SQ\pi = s(1), s(2), \dots, s(r)$; ako Q pripada $v(SQ\pi)$, onda je Q podniz od $SQ\pi$, ne novi niz.
3. Obnavlja se Q tako da bude $s(r+1), s(r+2)$, i ocjenjuje pripada li Q $v(SQ\pi)$ ili ne.
4. Ponavljaju se prethodni koraci dok Q ne pripada $v(SQ\pi)$. Sada $Q = s(r+1), s(r+2), \dots, s(r+i)$ nije podniz od $SQ\pi = s(1), s(2), \dots, s(r+i-1)$, pa se povećava $c(n)$ za jedan.
5. Nakon tog, S se obnavlja tako da bude $S = s(1), s(2), \dots, s(r+i)$, a $Q = s(r+i+1)$.

Ovaj postupak se ponavlja dok Q ne postane zadnji znak. U ovom trenutku broj različitih podnizova u P – mjera složenosti, je $c(n)$.

Da bi se dobila mjera složenosti neovisna o duljini niza, $c(n)$ bi se trebao normalizirati. Ako je duljina niza n i broj različitih simbola u skupu simbola je α , gornja granica od $c(n)$ dobiva se kao:

$$c(n) < \frac{n}{(1 - \varepsilon_n) \log_\alpha(n)}, \quad (18)$$

gdje je ε_n mala veličina i $\varepsilon_n \rightarrow 0$ ($n \rightarrow \infty$). Općenito, $n/\log_\alpha(n)$ je gornja granica od $c(n)$, gdje je baza logaritma α , odnosno:

$$\lim_{n \rightarrow \infty} c(n) = b(n) \equiv \frac{n}{\log_\alpha(n)} \quad (19)$$

Za pretvorbu 0-1 $\alpha = 2$, te slijedi

$$b(n) \equiv \frac{n}{\log_2(n)}, \quad (20)$$

a $c(n)$ se može normalizirati preko $b(n)$

$$C(n) = \frac{c(n)}{b(n)} \quad (21)$$

$C(n)$ odražava rastuću stopu novih obrazaca, odnosno bilježi vremensku strukturu slijeda.

3.3.8. Pogreška nelinearnog predviđanja (engl. *nonlinear prediction, NLP*)

Algoritam za dobivanje pogreške nelinearnog predviđanja prikazan je u nastavku [22]:

1. Skup podataka veličine n podijeljen je u dva dijela jednake veličine. Prvi dio ($y_1, y_2, \dots, y_{n/2}$) se koristi kao skup za učenje za izgradnju modela dinamike sustava, a drugi ($x_1, x_2, \dots, x_{n/2}$) se koristi kao skup za testiranje, koji se uspoređuje s predviđenim nizom.
2. Za dimenziju ugrađivanja m i vremenski odmak τ rekonstruira se atraktor svakog signala primjenom Takensovog teorema ugrađivanja (vremenski odgođeno ugrađivanje).
3. Za računanje prediktora X_{t+T}^* , gdje je T vrijeme za koje se vrši predviđanje, u skupu za učenje korištenjem Euklidske udaljenosti pronalazi se $m+1$ najbližih susjeda (Y_k) od X_t .
4. Svakom susjedu se pridružuje težina W_k ($k = 1, 2, \dots, m+1$), ovisno o njegovoj originalnoj udaljenosti

$$W_k = \frac{\|X_t, Y_k\|^{-2}}{\sum_{j=1}^{m+1} \|X_t, Y_j\|^{-2}}, \quad (22)$$

gdje je $\| \dots \|$ Euklidska udaljenost.

5. Predviđena se vrijednost dobiva računanjem gdje se susjedi pomiču T koraka u budućnost:

$$X_{t+T}^* = \sum_{k=1}^{m+1} W_k Y_{k+T}, \quad (23)$$

gdje je Y_{k+T} vrijednost k -tog susjeda T koraka u budućnosti.

6. Zatim se računa razlika predviđenog i stvarnog vektora, čime se dobiva vektor koji predstavlja pogrešku.

7. Kako bi se normalizirala pogreška predviđanja, računa se slučajna pogreška predviđanja prema izrazu (23), uz razliku da se umjesto susjeda uzima $m+1$ slučajnih točaka, a njihove se težine računaju u odnosu na neku slučajno odabranu točku.
8. Posljednji se korak sastoji od računanja srednjih kvadratnih vrijednosti za pogrešku predviđanja i slučajnu pogrešku predviđanja, a konačan rezultat dan je omjerom te dvije vrijednosti.

3.3.9. Zajednička dimenzija (engl. *mutual dimension*, D_m)

Za računanje D_m za dva vremenska niza A i B potrebno je najprije izračunati korelacijsku dimenziju za niz A (D_2A) i korelacijsku dimenziju za niz B (D_2B), obje za maksimalnu dimenziju ugrađivanja M . Zatim se promatra ujedineni fazni prostor od ukupno $2 \times M$ dimenzija. U ovom prostoru konstruira se zajednička trajektorija C. Svaka točka iz C sastoji se od $2 \times M$ koordinata, M dobivenih iz A i M dobivenih iz B. Računa se korelacijska dimenzija za C (D_2C). Zajednička dimenzija definirana je kao

$$D_m = D_2A + D_2B - D_2C \quad (24)$$

Ako oba vremenska niza A i B odgovaraju istom atraktoru, vrijedit će $D_2A = D_2B = D_2C$. U tom slučaju zajednička će dimenzija odgovarati dimenziji atraktora svakog od nizova, odnosno M . Ako oba vremenska niza imaju različite atraktore (njihova dinamika je nepovezana), očekuje se $D_2A + D_2B = D_2C$ [35].

3.3.10. Značajke cross-recurrence plota (CRP)

CRP je bivarijatno proširenje recurrence plot, a njime se analiziraju ovisnosti između dva različita sustava uspoređivanjem njihovih stanja [25]. Za dva dinamička sustava s trajektorijama $\vec{x}_i$ i $\vec{y}_j$ u d -dimenzionalnom prostoru stanja odgovarajuća cross recurrence matrica definirana je kao

$$CR_{i,j}^{\vec{x},\vec{y}}(\varepsilon) = \theta(\varepsilon - \|\vec{x}_i - \vec{y}_j\|), \quad (25)$$

za $i = 1, \dots, N$, $j = 1, \dots, M$.

3.3.10.1. CRP recurrence rate (CC)

Ova se značajka temelji na kros-korelacijskoj sumi (engl. *cross-correlation sum*) definiranoj izrazom:

$$C_{I,J}(\varepsilon) = \frac{1}{w^2} \sum_{i,j=1}^w R_{i+(I-1)w, j+(J-1)w}(\varepsilon), \quad (26)$$

za $I, J = 1, \dots, \frac{N}{w}$

Definirana je kao:

$$CC_2(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N CR_{i,j}(\varepsilon), \quad (27)$$

3.3.10.2. CRP determinizam (DET)

Ova se značajka temelji na dijagonalnim linijama i predstavlja mjeru predvidljivosti sustava. Definirana je kao:

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=1}^N lP(l)}, \quad (28)$$

gdje je $P(l)$ dijagonalnih linija duljine l ($P(l) = P(\varepsilon, l)$) definirana kao

$$P(\varepsilon, l) = \sum_{i,j=1}^N \left(1 - R_{i-1,j-1}(\varepsilon)\right) \left(1 - R_{i+1,j+1}(\varepsilon)\right) \prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon) \quad (29)$$

3.3.10.3. CRP laminarity (LAM)

Značajka se temelji na vertikalnim linijama, a računa se sljedećom formulom:

$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=1}^N vP(v)}, \quad (30)$$

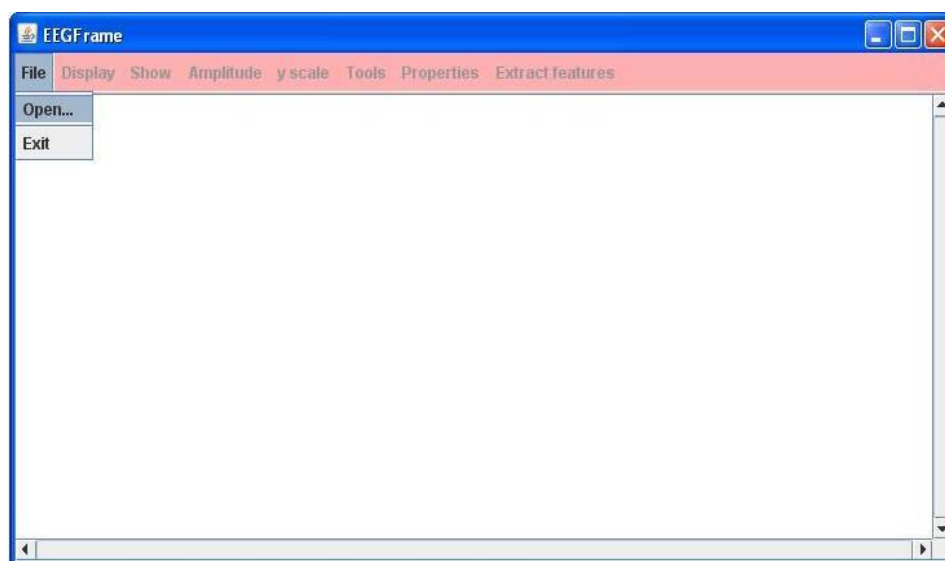
gdje je v duljina vertikalnih linija.

4. Opis razvijenog radnog okvira i primjeri korištenja

Računalni radni okvir, nazvan *EEGFrame*, izrađen je u programskom jeziku Java, što omogućava prenosivost, smanjenje složenosti zahvaljujući objektno usmjerenoj arhitekturi, te olakšanu izradu grafičkog korisničkog sučelja. U izradi je korišteno integrirano razvojno okruženje Eclipse IDE za Javu. Nastao je usavršavanjem i nadogradnjom ranije razvijenog programskog produkta za prikaz elektroencefalografskih signala [37], čija je funkcionalnost proširena uvođenjem izlučivanja linearnih i nelinearnih značajki s ciljem modeliranja i predviđanja poremećaja. Pritom je prilagođen kôd radnog okvira za analizu varijabilnosti srčanog ritma (*HRVFrame* [38]) za korištenje u *EEGFrameu*. U *HRVFrameu* je implementirano više od 30 značajki linearne vremenske, frekvencijske i vremensko-frekvencijske domene, te nelinearne značajke, a mnoge od njih se mogu primijeniti na analizu drugih biomedicinskih vremenskih nizova. Istraženo je koje se značajke iz *HRVFramea* koriste u analizi elektroencefalografskih signala, one su prilagođene na odgovarajući način i dodane u *EEGFrame*, a implementirane su i neke značajke specifične za EEG, uz provjeru implementacije relevantnih metoda na temelju izvorne literature. Značajke su organizirane u paketima s odgovarajućim imenom. Primjerice, u paketu *features.nonlinear.entropy* nalaze se postupci za izlučivanje značajki koje se odnose na entropiju. *EEGFrame* ima mogućnost učitavanja EEG signala iz datoteke u EDF formatu nastale korištenjem sustava 10-20 za raspored elektroda i nomenklaturu. Omogućeno je prikazivanje više signala odjednom na različitim vremenskim skalama s različitim parametrima za prikaz, spremanje uzoraka odabranih signala u tekstualnu datoteku, prikaz relevantnih metapodataka koji se nalaze u zaglavlju učitane datoteke i prikaz svojstava nekog od signala. Osim tog, ostvaren je ispis izlučenih značajki u izlaznu datoteku u CSV formatu. Programski produkt će se u budućnosti nadopunjavati novim značajkama i opcijama. Trenutna verzija *EEGFramea* dostupna je na adresi: <http://www.zemris.fer.hr/~ajovic/eegframe/eegframe.html>.

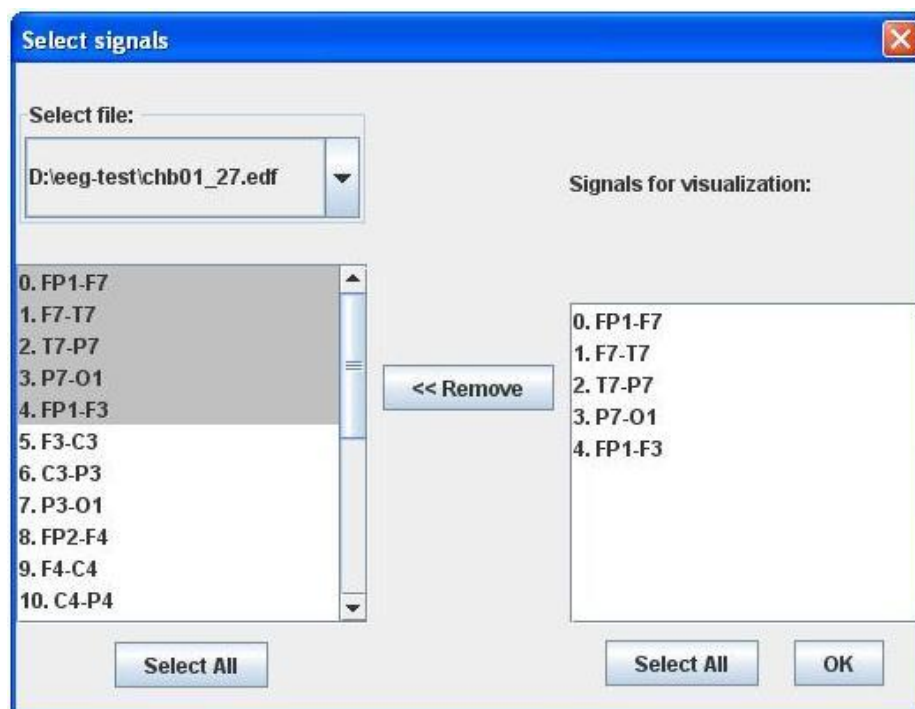
4.1. Mogućnosti i primjeri korištenja EEGFramea

Testiranje se odvijalo na EDF datotekama iz referentne internetske banke fizioloških podataka PhysioBank [39]. Mogućnosti i primjeri korištenja razvijenog radnog okvira prikazane su na datotekama iz baze „CHB-MIT Scalp EEG Database“. Sve datoteke su u osnovnom EDF formatu, nastale korištenjem sustava 10-20. U nastavku će se rad demonstrirati na datoteci chb01_27.edf, koja se sastoji od 23 signala mjerenih bipolarno. Prilikom pokretanja otvara se glavni prozor koji se sastoji od izbornika i pozadine za prikaz signala (Slika 8). U početku je moguće odabrati samo opciju *File* u izborniku. Klikom na *Open...* otvara se dijalog za odabir ulazne datoteke, koja mora biti u osnovnom formatu opisanom u poglavlju „2.4. EDF format“, ili u proširenom, EDF+ formatu. U primjeru je odabrana datoteka chb01_27.edf.



Slika 8. Pokretanje – početni prozor

Nakon odabira ulazne datoteke i učitavanja podataka, pojavljuje se dijalog za odabir signala za prikaz (Slika 9). Isti dijalog se pojavljuje i odabirom opcije *Display* → *Add/Remove Signals*.

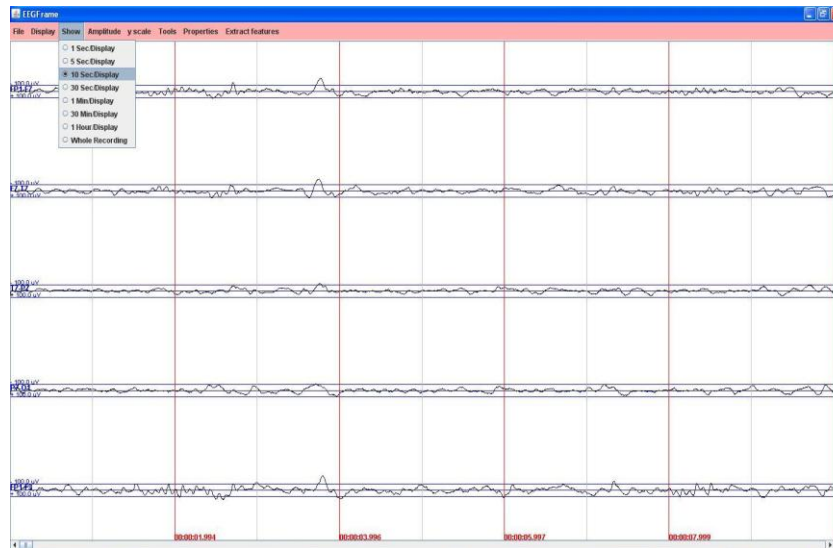


Slika 9. Dijalog za odabir signala za prikaz ili brisanje. Na slici je odabrano pet signala, koji se iscrstavaju pritiskom na tipku OK

4.1.1. Opcije za prikaz signala

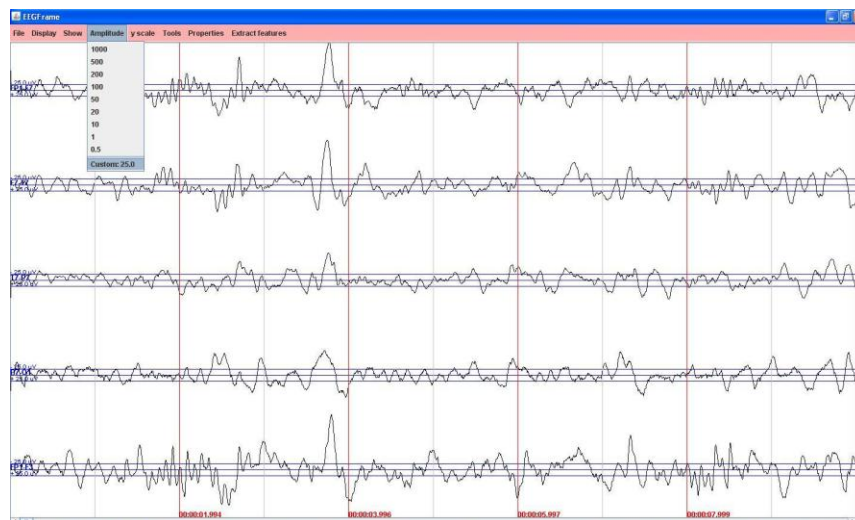
Postoji mogućnost odabira različitih opcija za prikaz signala. Odabirom opcije *Show* u izborniku signali mogu biti prikazani na različitim vremenskim skalama.

Primjerice, odabirom opcije *10 Sec/Display* (Slika 10), po jednom je prikazu vidljivo deset sekundi signala, a svakih deset sekundi iscrtana je vertikalna crvena linija u svrhu bolje preglednosti. Osim toga, između su u određenim trenucima iscrtane pomoćne vertikalne crvene linije sa zapisom vremena proteklog od početka datoteke, kako bi se korisnik mogao lakše snaći. Deset sekundi po prikazu je ujedno i zadana opcija, koja je automatski odabrana kod prvog odabira signala za prikaz.



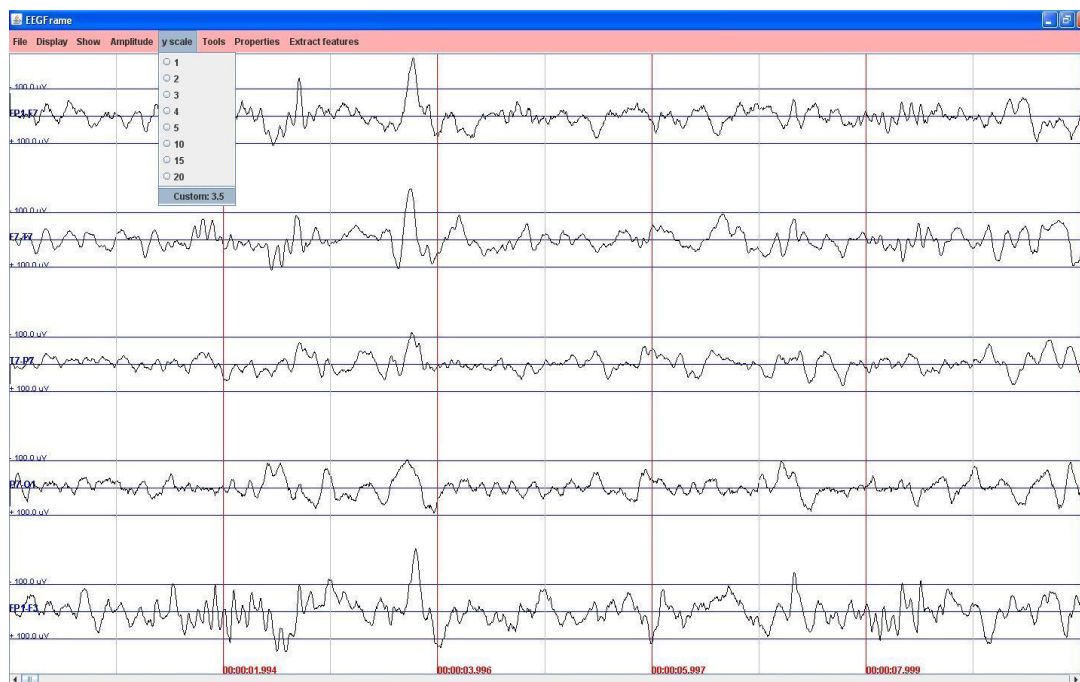
Slika 10. Prikaz pet odabranih signala na vremenskoj skali od 10 sekundi po prikazu s amplitudom 100 μV , na dnu su u određenim trenucima prikazane vremenske oznake

Odabirom opcije *Amplitude* u izborniku moguće je mijenjati amplitudu signala. Amplituda je u EDF datoteci najčešće zadana u mikrovoltima (μV), a kako je amplituda spontanog EEG-a reda veličine 100 μV , zadana amplituda kod prvog odabira signala za prikaz iznosi 100. Ako je prikazano više signala odjednom, svi imaju odabranu jednaku amplitudu. Odabirom opcije *Custom* u izborniku *Amplitude* korisnik može unijeti proizvoljnu vrijednost amplitude, ukoliko mu nijedna od ponuđenih opcija ne odgovara (Slika 11).



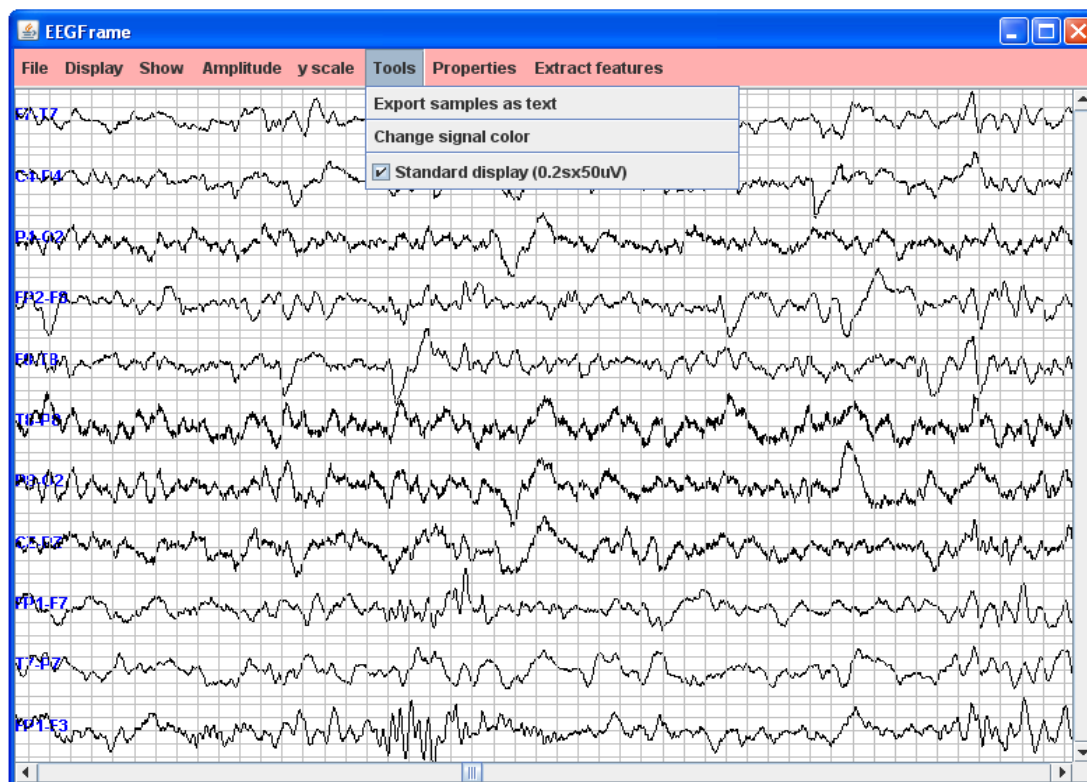
Slika 11. Prikaz pet odabranih signala na vremenskoj skali od 10 sekundi po prikazu, uz proizvoljno unesenu amplitudu (iznosa 25)

Kod prikaza signala postoji i mogućnost vertikalnog skaliranja, odabirom opcije *y scale* u izborniku (Slika 12). Zadana vrijednost faktora vertikalnog skaliranja je 1, i odabrana je kod prvog dodavanja ili micanja signala. Odabirom opcije *Custom* korisnik ima više slobode i mogućnost detaljnijeg prikaza signala, jer sam unosi željenu vrijednost.



Slika 12. Prikaz pet odabranih signala na vremenskoj skali od 10 sekundi po prikazu, s amplitudom 100 μV , uz proizvoljno unesen faktor skaliranja (iznosa 3.5)

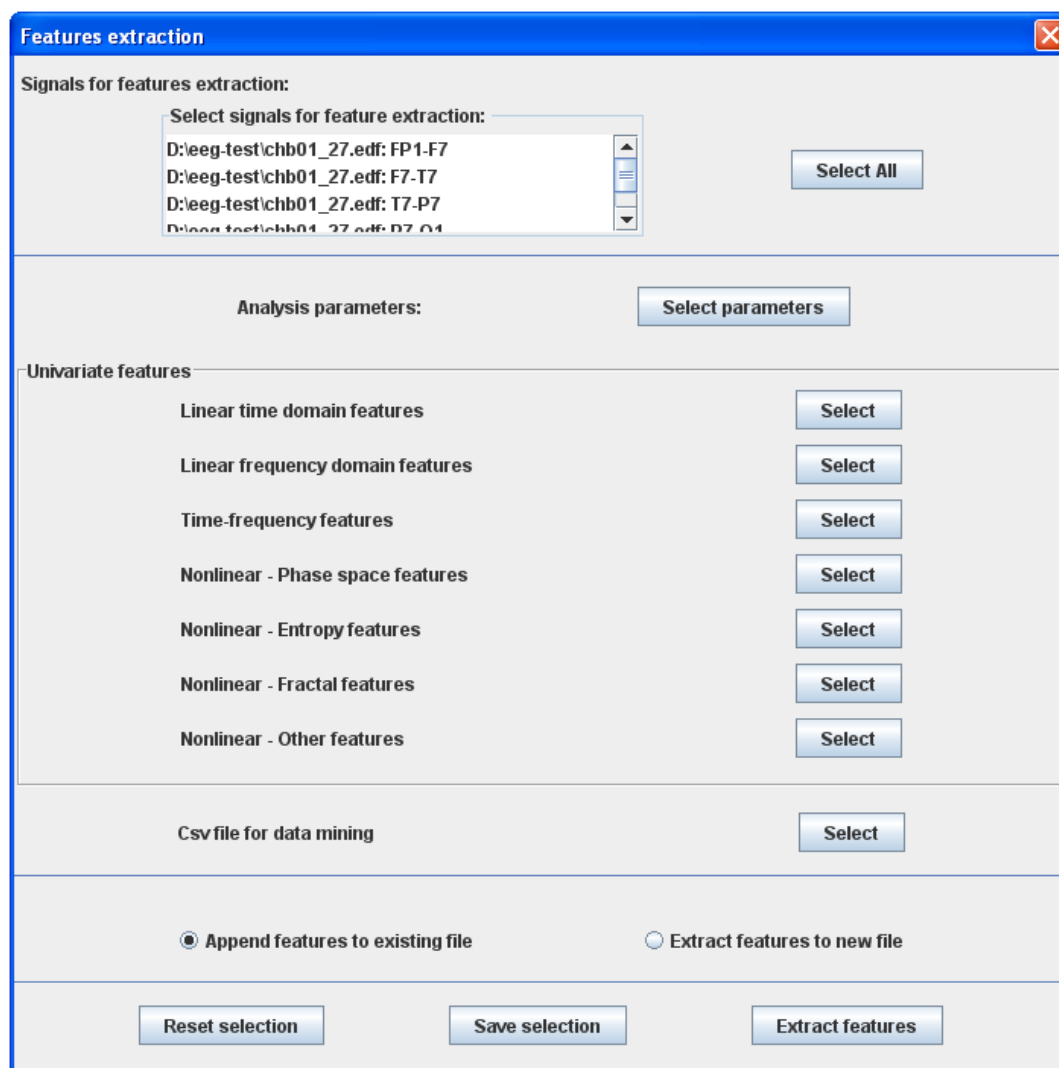
Signale je moguće prikazati pomoću standardne opcije 0.2 s x 50 μV . Pritom razmak između dvije uzastopne vertikalne linije predstavlja vremenski interval od 0.2 sekunde, a razmak između dvije horizontalne linije čini 50 μV . Pritom su svi signali prikazani unutar istog prozora, bez mogućnosti vertikalnog pomicanja. Ova opcija se odabire označavanjem *Tools* → *Standard display(0.2sx50uV)* (Slika 13).



Slika 13. Opcija 0.2 s x 50 μ V uz odabranih 10 s po prikazu – svi signali su prikazani u istom prozoru, razmak između dvije vertikalne linije predstavlja 0.2 s, a razmak između dvije horizontalne linije predstavlja 50 μ V

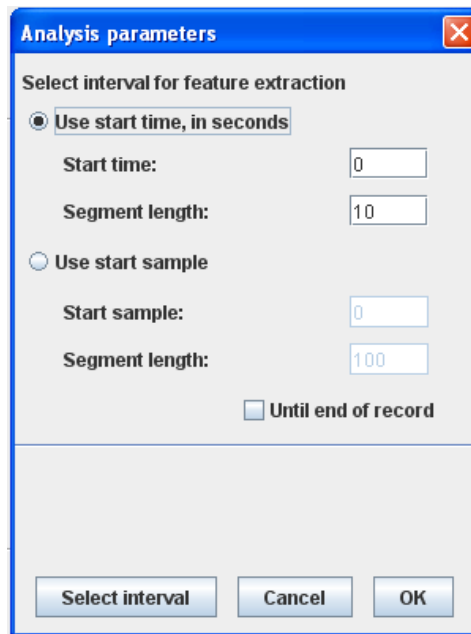
4.1.2. Funkcionalnost za izlučivanje značajki

Značajke se mogu izlučiti odabirom opcije *Extract features* u izborniku, te potom *Univariate features*, odnosno *Multivariate features*, ovisno o željenom tipu značajki. Dijalog koji se otvara odabirom opcije *Univariate features* prikazan je na slici 14.



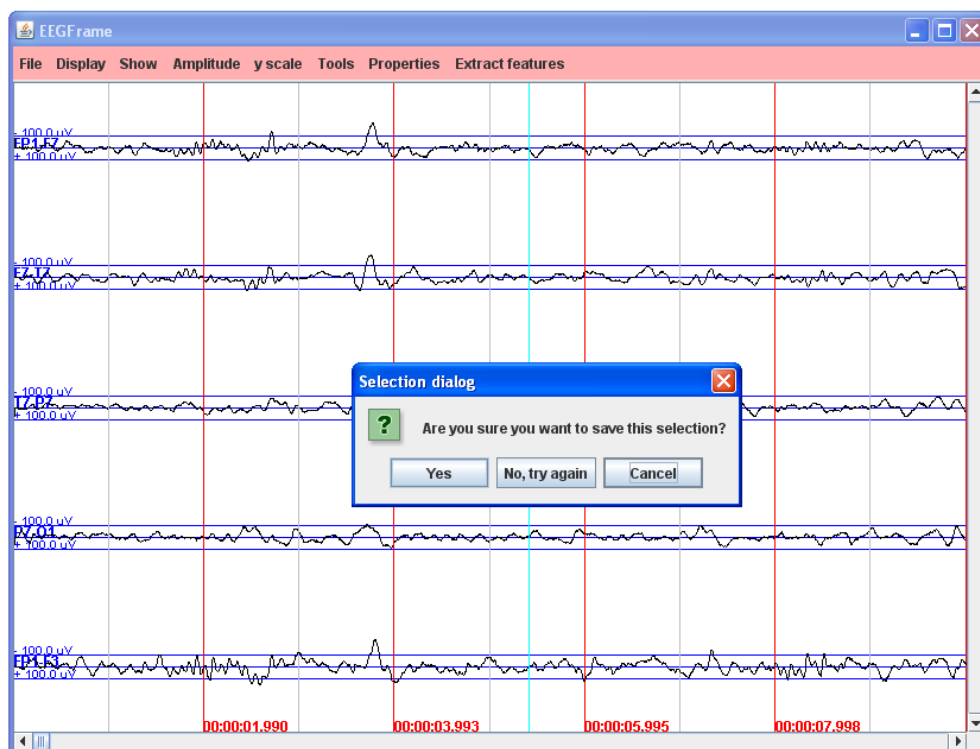
Slika 14. Dijalog za izlučivanje univarijatnih značajki

Na vrhu se nalaze signali koje je moguće odabrati za postupak izlučivanja značajki. Sve odabrane univarijatne značajke izlučit će se na svakom od ovih signala za odabrani vremenski interval. Ispod signala se nalazi opcija *Analysis parameters*, kojom je, pritiskom na *Select parameters* moguće podesiti interval za koji će se izvršiti ekstrakcija. Interval se može izabrati u sekundama, pri čemu se unosi početno vrijeme i veličina segmenta za koji se provodi analiza u sekundama (unaprijed zadane opcije su redom 0 i 10), ili u broju uzoraka, gdje se unosi redni broj početnog uzorka, te veličina segmenta u broju uzoraka (unaprijed zadane opcije su redom 0 i 100). Ako je označena opcija *Until end of record*, analiza će se provesti od unesenog početnog vremena ili uzorka sve do kraja zapisa, a parametar veličina segmenta će se ignorirati (Slika 15).



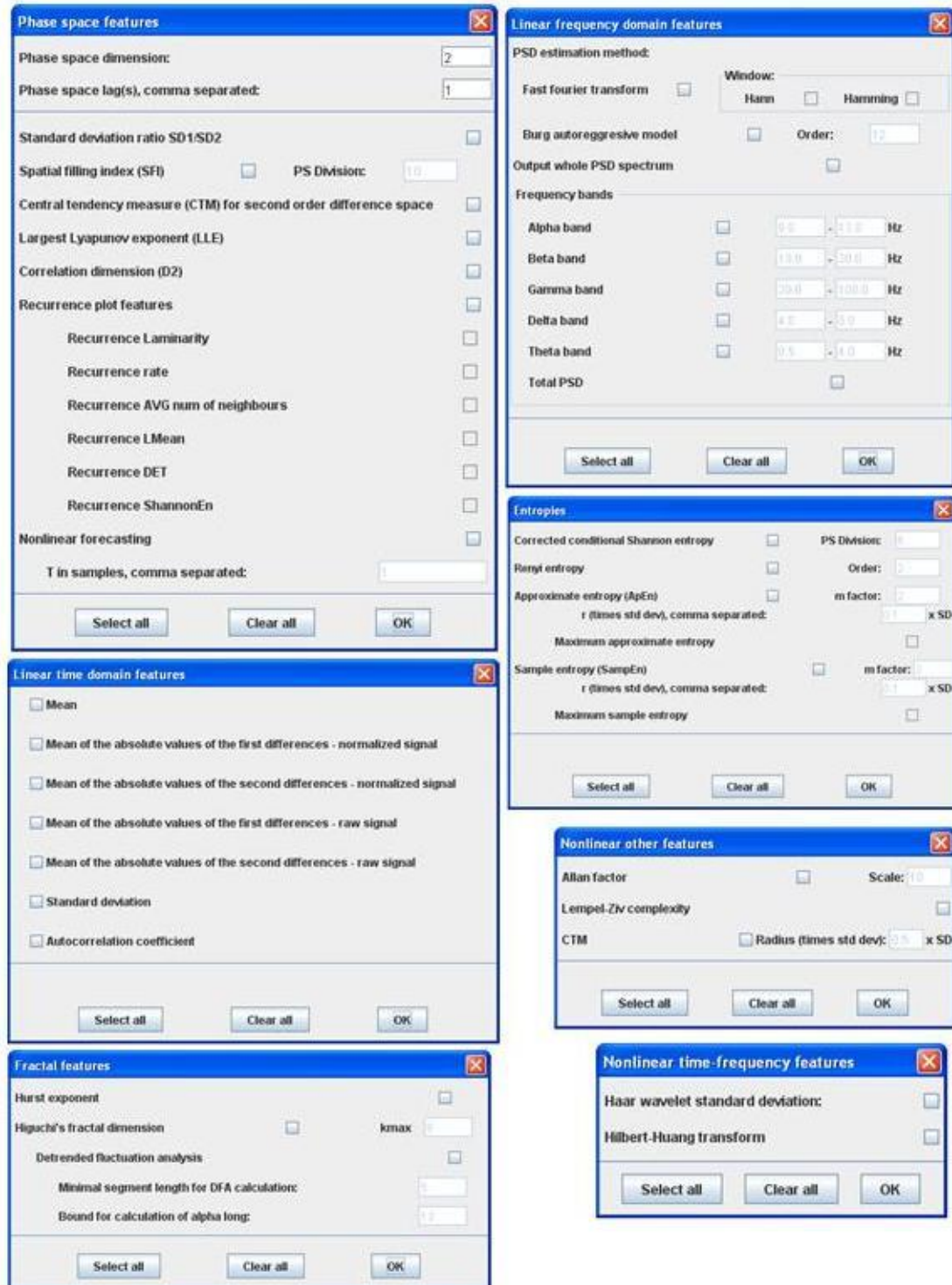
Slika 15. Dijalog za odabir segmenta

Osim unosa početka, odnosno trajanja segmenta pomoću dijaloga *Analysis parameters*, korisnik ima mogućnost označiti željeni interval izravno na pozadini za prikaz signala, tako da u glavnom izborniku odabere *Extract features* → *Select interval* → *Select start/end point*, i potom klikne na željenu x koordinatu na pozadini, koja predstavlja trenutak koji će se pohraniti umjesto ranije odabranih vrijednosti u *Analysis parameters* dijalogu. Pritom će se kroz odabranu x koordinatu iscrtati vertikalna linija svijetloplave boje i pojaviti će se dijalog kojim je moguće spremiti odabranu vrijednost, pokušati ponovno ili odustati (Slika 16).



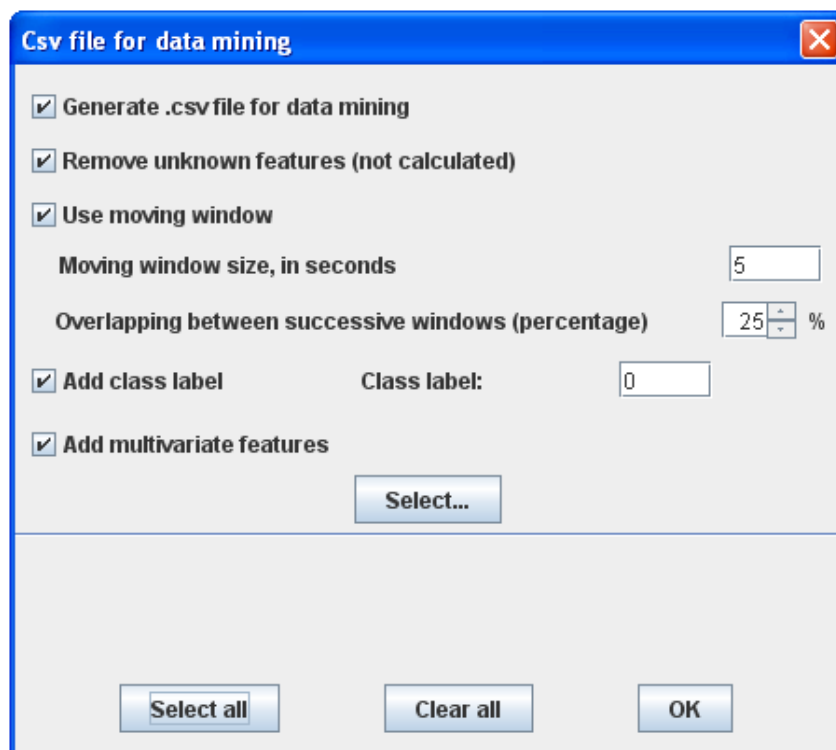
Slika 16. Označavanje intervala izravno na vizualiziranim signalima

Ispod parametara za analizu nalazi se panel na kojem je moguće odabrati značajke podijeljene na linearne vremenske, frekventijske, vremensko-frekventijske, nelinearne značajke faznog prostora, nelinearne značajke temeljene na entropiji, fraktalne i ostale nelinearne značajke. Pritiskom na *Select* za pojedinu klasu značajki otvara se dijalog s pripadajućim značajkama i njihovim parametrima. Svi dijalozi prikazani su na slici 17.



Slika 17. Prikaz dijaloga za izlučivanje značajki

Opcija *Csv file for data mining* s mnogim korisnim opcijama omogućava jednostavno izlučivanje značajki i ispis u datoteku s ekstenzijom „csv“ (značajke su međusobno odvojene zarezima) pogodnu za dubinsko pretraživanje podataka (Slika 18).



Slika 18. Dijalog koji se otvara pritiskom na *Select* za opciju *Csv file for data mining*

Ako nije označena opcija *Generate .csv file for data mining* u ovom dijalogu, značajke će se u izlaznoj datoteci ispisati u poseban redak za svaki odabrani signal, bit će ispisane sve značajke (one koje nisu izračunate bit će označene upitnikom) a izlučit će se samo za jedan vremenski segment, onaj zabilježen u *Analysis parameters*. Označavanjem ove opcije značajke svih signala za jedan vremenski segment konkatenerane su tako da čine veliki vektor značajki ispisan u jednom redu. Označavanjem *Remove unknown features* u izlaznoj se datoteci uklanjaju značajke koje nisu odabrane za računanje. Omogućeno je korištenje pomičnog prozora, čime se vremenski segment iz *Analysis parameters* dijeli na više manjih segmenata, od kojih je svaki ima duljinu navedenu u *Moving window size, in seconds*. Pritom se uzastopni intervali (susjedni prozori) mogu djelomično preklapati, što se zadaje postotkom preklapanja. Za svaki prozor, odnosno vremenski interval, stvara se novi red u

datoteci. Označavanjem opcije *Add class label*, u izlaznu se datoteku dodaje stupac s nazivom *Class*, koji je koristan prilikom klasifikacije. Zadana oznaka je „0“, no kao oznaku razreda moguće je upisati proizvoljan niz znakova. Označavanjem opcije *Add multivariate features* omogućeno je računanje multivarijatnih značajki uz univarijatne, te se one u izlaznoj datoteci jednostavno nadovezuju na vektor univarijatnih značajki, povećavajući mu time dimenzionalnost. Odabirom opcije *Select...* otvara se dijalog s multivarijatnim značajkama prikazan na slici 19.

Multivariate nonlinear features

Phase space dimension 1: 2

Phase space lag(s), comma separated: 1

Phase space dimension 2: 2

Phase space lag(s), comma separated: 1

Mutual dimension ☒

Finesse: 20

Cross recurrence plot features ☐

Cross recurrence laminarity ☐

Cross recurrence rate ☐

Cross recurrence LMean ☐

Cross recurrence DET ☐

Cross recurrence ShannonEn ☐

Select signals:

- FP1-F7_F7-T7
- FP1-F7_T7-P7
- FP1-F7_P7-O1
- FP1-F7_FP1-F3
- F7-T7_T7-P7

Select All

Select all Clear all OK

Slika 19. Dijalog koji omogućava odabir multivarijatnih značajki i njihovih parametara. Signali za odabir nastali su različitim kombinacijama signala odabranih za prikaz

Prije samog izlučivanja značajki moguće je odabrati hoće li se stvoriti nova datoteka u koju će se ispisati značajke (*Extract features to new file*), ili će se one nadovezati na sadržaj postojeće datoteke (*Append features to existing file*).

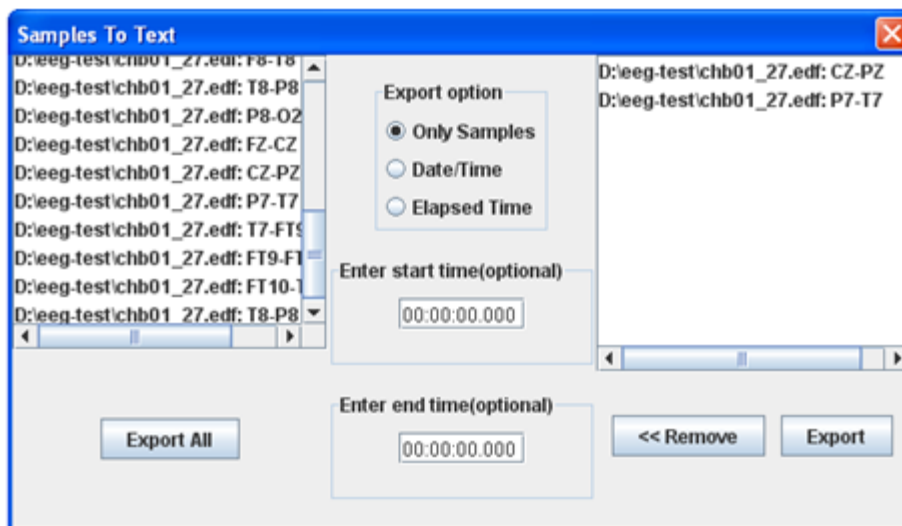
Ukoliko se na početku odabere *Extract features* → *Multivariate features*, otvara se dijalog vrlo sličan dijalogu za univarijatne značajke, a postupak odabira značajki analogan je postupku opisanom za univarijatne značajke, te se neće posebno obraditi.

4.1.3. Dodatne opcije

Razvijeni radni okvir uz vizualizaciju signala i izlučivanje značajki ima i mogućnost ispisa uzoraka odabranih signala u tekstualnu datoteku, te prikaza bitnih informacija iz zaglavlja datoteke.

4.1.3.1. Ispis uzoraka u tekstualnu datoteku

Odabirom opcije *Tools* → *Export Samples As Text* otvara se dijalog u kojem je moguće izabrati željene signale za ispis, vrstu ispisa (samo uzorke, uzorke i proteklo vrijeme, uzorke i stvarno vrijeme i datum koji su protekli od početka snimanja), kao i ispis od željenog početnog do željenog završnog trenutka (Slika 20). Ukoliko korisnik ne upiše početno i završno vrijeme, ispisat će se svi uzorci, bez obzira na odabranu vrstu ispisa. Opcija *Only Samples* ispisuje uzorke i najmanji vremenski razmak između dva uzorka (u slučaju različitih frekvencija uzorkovanja). Opcija *Date/Time* ispisuje vrijeme u formatu hh:mm:ss.mmm, te datum u formatu dd/mm/yyyy, a prilikom unosa početnog ili završnog vremena potrebno je paziti na to da se vrijeme računa od stvarnog vremena početka snimanja.

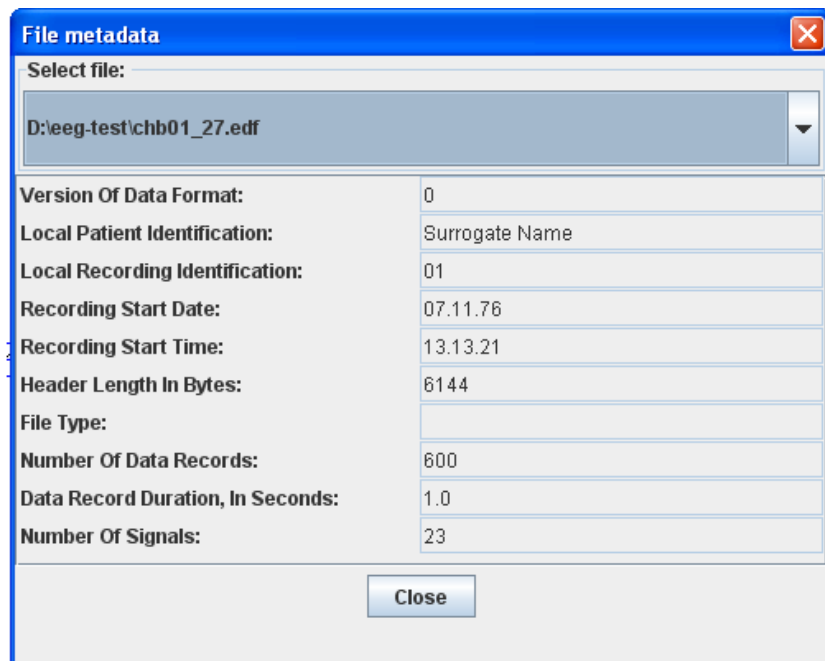


Slika 20. Dijalog za odabir signala i opcija za ispis u tekstualnu datoteku

Opcija *Elapsed Time* omogućuje ispis proteklog vremena, a kao i *Only Samples* opcija, ne podržava korisnički unos koji ovisi o stvarnom vremenu početka snimanja. U slučaju da korisnik unese početno vrijeme veće od trajanja čitavog zapisa, u datoteku će se ispisati samo nazivi signala. Ako je unesen početni trenutak manji od početnog trenutka prvog uzorka, ispis će početi od prvog uzorka. Ako je unesen završni trenutak veći od završnog trenutka zadnjeg uzorka, ili je završni trenutak manji od početnog, ispisat će se svi uzorci, te će ispis stati na posljednjem.

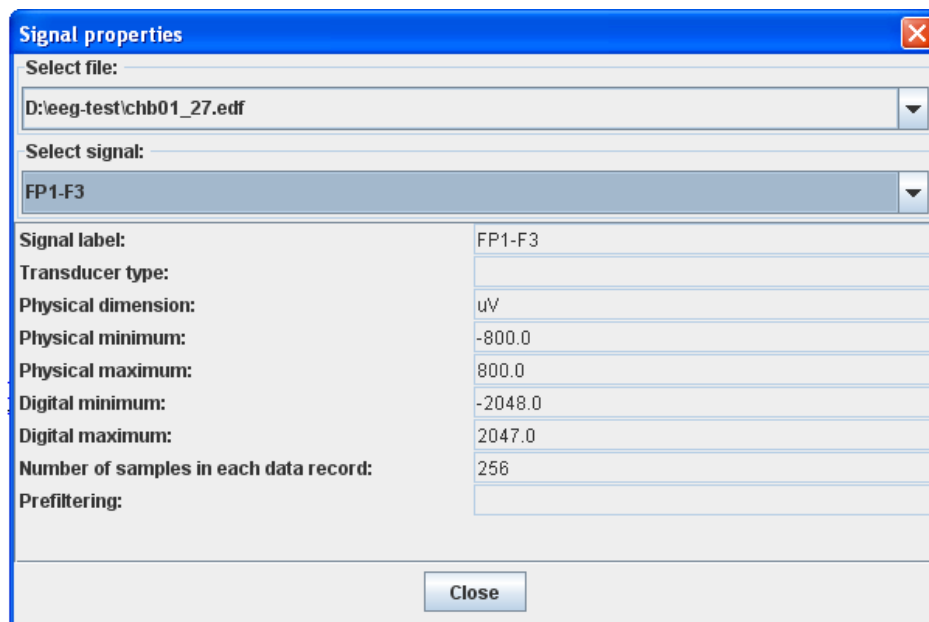
4.1.3.2. Prikaz informacija iz zaglavlja

Moguće je izabrati prikaz podataka o ulaznoj datoteci i signalima, koji su zapisani u njezinom zaglavlju. Odabirom opcije *Properties* → *EDF file* otvara se prozor koji prikazuje informacije o datoteci i snimanju zapisane u zaglavlju, kao što su verzija, broj zapisa podataka, trajanje zapisa podataka, broj signala u svakom zapisu i drugi (Slika 21).



Slika 21. Prikaz informacija o datoteci i snimanju

Odabirom opcije *Properties* → *Signals* otvara se prozor sa signalima, a odabirom željenog signala prikazuju se podaci o njemu navedeni u zaglavlju datoteke, poput fizikalne veličine, broja uzoraka u svakom zapisu podataka i drugo (Slika 22).



Slika 22. Prikazan je prozor sa svojstvima signala; odabran je signal FP1-F3

5. Analiza epilepsije

5.1. Dinamička priroda epileptičkih napadaja

Epileptički su napadaji, za razliku od normalne aktivnosti u pozadini, visoko nelinearni fenomeni [4]. EEG zapisi pacijenata s epilepsijom pokazuju dvije kategorije abnormalne aktivnosti: abnormalni signali snimljeni između epileptičkih napadaja (engl. *interictal*) i aktivnost zabilježena tijekom epileptičkog napadaja (engl. *ictal*).

Prva zabilježena nelinearna analiza epileptičkog napadaja zabilježila je bitno nižu korelacijsku dimenziju od normalnog EEG-a mirovanja, što je nagovijestilo da bi epileptički napadaji mogli nastati zbog patološkog gubitka složenosti. S tim konceptom se slaže u drugoj studiji zabilježeno smanjenje najvećeg Lyapunovljevog eksponenta tijekom epileptičkog napadaja [40]. U analiziranju podataka s prikladno konstruiranim zamjenskim podacima, zaključeno je da dinamika šiljaka i valnih ispuštanja nije kaotična ali odražava šumovit periodični atraktor, što je potvrđeno u brojnim kasnijim studijama. Mnoge studije koristile su neku vrstu testiranja zamjenskim podacima za istraživanje nelinearne prirode napadaja, a kao posljedica toga postoji prilično čvrst dokaz da napadaji snažno odražavaju nelinearnu dinamiku mozga.

Činjenica da je aktivnost napadaja visoko nelinearna i vjerojatno niskodimenzionalna, a EEG snimljen između epileptičkih napadaja je visokodimenzionalan i samo slabo nelinearan otvara pitanje kako se odvija tranzicija u aktivnost tijekom epileptičkog napadaja, a u vezi ovog pitanja predloženo je više mogućih scenarija [41].

Osim promjena u lokalnoj dinamici atraktora, napadaji mogu biti karakterizirani promjenama u povezanosti između različitih područja mozga. Ovdje treba uzeti u obzir da studije uključuju različite tipove uzoraka različitih pacijenata i različite mjere sinkronizacije, što je moglo utjecati na rezultate. Iako su napadaji uglavnom karakterizirani porastom u povezanosti između različitih područja mozga, postoje indikacije da u nekim tipovima napadaja zapravo postoji smanjenje u razini povezanosti koje prethodi napadaju [4].

5.2. Detekcija i predviđanje napadaja

Nakon ranih početaka rada na predvidljivosti napadaja 1970-ih, različite su grupe uz pomoć linearnih pristupa pokušale izlučiti preteče napadaja iz površinskih EEG zapisa. Pojavom fizikalno-matematičke teorije nelinearnih sustava u 1980-ima, uvedeni su novi pristupi koji su bili usmjereni na bolju karakterizaciju dinamičkih sustava otkrivajući složenije ponašanje nego što se do tada moglo prikazati s klasičnim linearnim pristupima [2]. Tijekom razdoblja ranog optimizma predloženo je više algoritama za nelinearno predviđanje epileptičkih napadaja, uključujući, između ostalog, korelacijske dimenzije, Lyapunovljeve eksponente i entropijske mjere. Drugi su se pristupi fokusirali na procjenu promjena u nelinearnoj povezanosti između različitih područja mozga pomoću fazne sinkronizacije.

Međutim, rezultati ranih istraživanja pokazali su se upitnima, jer nisu mogli biti reproducirani u kasnijim studijama, što je potaklo rast skepticizma. Za neke od njih zabilježena je smanjena sposobnost predviđanja, ili se ispostavilo da su se jednako dobri rezultati mogli dobiti pomoću linearnih postupaka. Postalo je jasno da daljnji napredak u ovom području ovisi o razvoju prikladnih statističkih testova [4].

Različite su skupine 2005. objavile niz studija provedenih na skupu od pet stalnih višednevnih zapisa dobivenih od različitih centara za epilepsiju za prvu „Međunarodnu kolaborativnu radionicu o predviđanju epileptičkih napadaja“ (engl. *First International Collaborative Workshop on Seizure Prediction*), održanu u Bonnu u travnju 2002. [2]. Cilj ove radionice bio je da različite skupine testiraju i usporede svoje metode na zajedničkom skupu podataka. Rezultati iz različitih skupina uglavnom su pokazali lošu izvedbu univarijatnih mjera, dok su bolje performanse zabilježene za bivarijatne i multivarijatne značajke, iako je utvrđeno da su promatrane promjene prije epileptičkog napadaja lokalno ograničene na specifične lokacije snimanja. Jedna od studija sadržala je opsežnu usporedbu izvedbe predviđanja niza univarijatnih i bivarijatnih značajki, obuhvativši i linearne i nelinearne pristupe, koristeći koncept zamjenskih podataka za procjenu statističke značajnosti dobivenih rezultata [42]. Autori su otkrili značajne prediktivne performanse za mjere sinkronizacije, dok univarijatne mjere, uključujući i korelacijsku dimenziju, najveći

Lyapunovljev eksponent i energiju signala, nisu mogli razlikovati aktivnost tijekom epileptičkog napadaja od aktivnosti između epileptičkog napadaja. Nelinearne mjere nisu pokazale veće prediktivne performanse od linearnih mjera [2].

Korištenjem komparativnih testova i statističkih kontrolnih postupaka, može se reći da će daljnji napredak u predviđanju i detekciji napadaja biti moguć tijekom idućih nekoliko godina [4].

5.3. Primjena EEGFramea za analizu epilepsije

Izlučivanje značajki provedeno je na EDF datotekama iz baze „CHB-MIT Scalp EEG Database“, koje predstavljaju pedijatrijske EEG zapise. Pedijatrijski zapisi pokazuju veću varijabilnost u aktivnosti tijekom i između napadaja od EEG-a odraslih. Svaki slučaj sadrži između 9 i 42 kontinuirane .edf datoteke od jednog subjekta. U većini slučajeva .edf datoteke sadrže točno jedan sat digitaliziranih EEG signala, iako su zapisi koji pripadaju slučaju chb10 trajanja 2 sata, a oni koji pripadaju slučajevima chb04, chb06, chb07 i chb09 su trajanja 4 sata. Svi signali su uzorkovani frekvencijom od 256 Hz. Većina datoteka sadrži 23 bipolarno snimljena EEG signala. Prilikom snimanja korišten je međunarodni sustav 10-20 za raspored elektroda i nomenklaturu.

5.3.1. Cilj i opis provedbe analize

U radu je odabrana kombinacija više linearnih i nelinearnih univarijatnih i bivarijatnih značajki. Korištene su linearne vremenske značajke (srednja vrijednost), nelinearne fraktalne značajke (Hurstov eksponent), nelinearne značajke u faznom prostoru (korelacijska dimenzija, najveći Lyapunovljev eksponent, NLP), nelinearne značajke zasnovane na entropiji (SampEn), ostale nelinearne univarijatne značajke (Lempel-Ziv, CTM), nelinearne bivarijatne značajke (zajednička dimenzija, CRP značajke: CRP recurrence rate, CRP DET, CRP LAM).

Za značajke u faznom prostoru odabrani su parametri $l = 3$ za vremenski razmak, $m = 12$ za dimenziju ugrađivanja. SampEn je izračunat za $m = 1$ i četiri vrijednosti za r (0.1xSD, 0.15xSD, 0.2xSD, 0.25xSD, gdje je SD standardna devijacija originalnog vremenskog niza). CTM je izračunat za vrijednost 2xSD, gdje je SD standardna devijacija originalnog vremenskog niza. Greška nelinearnog predviđanja izračunata je za broj uzoraka $T = 1, 2, 3, 4$ i 5.

Cilj rada je binarna klasifikacija vektora značajki, ovisno o tome predstavlja li vektor značajki EEG zapis napadaja ili EEG zapis kada nema napadaja (tzv. *non-seizure* EEG). Iz određenih zapisa odabranih pacijenata izlučene su navedene značajke, tako da su pronađeni dijelovi zapisa gdje nije bilo poremećaja (barem 5 minuta prije početka poremećaja) i dijelovi zapisa gdje je postojao poremećaj. Kod uzimanja dijelova zapisa bez poremećaja, periodi unutar 30 minuta nakon početka poremećaja nisu uključeni u analizu, kako bi se isključili efekti perioda neposredno nakon poremećaja, u kojem su moguće izmjene u EEG-u. Ti intervali određeni su na temelju metapodataka o napadaju (vrijeme početka i završetka), koji su pohranjeni u tekstualnoj datoteci za svakog pacijenta u bazi. Odabrana je približno ista duljina normalnih i abnormalnih dijelova.

Zbog velike količine podataka, u svakom zapisu analizirano je samo nekoliko unaprijed odabranih signala. Za univarijatne značajke to su F3-C3, F4-C4, P3-O1, P4-O2, T8-P8, C3-P3, C4-P4. Izlučivanje bivarijantnih značajki ograničeno je na kombinacije nekih susjednih kanala. Odabrane su kombinacije F3-C3 i C3-P3, F7-T7 i T7-P7, F4-C4 i C4-P4, F8-T8 i T8-P8, FP1-F3 i F3-C3. Kako su EEG signali nestacionarni, značajke su izlučene iz razumno malog vremenskog intervala. Nije ih moguće podijeliti na kratke fiziološki relevantne segmente, pa se često koriste intervali duljine 2-5 sekundi [43]. Segmenti su u ovom radu dobiveni pomoću pomičnog prozora (engl. *sliding window*) veličine 3 sekunde, pri čemu je odabrano preklapanje od 25 % između susjednih prozora, čime je dobiveno više primjera za učenje iz jedne snimke. Na svim odabranim signalima izračunate su značajke na istom vremenskom segmentu, a za daljnju analizu generirani su vektori značajki na način da su vektori značajki iz različitih signala konkatenirani tako da formiraju veliki vektor značajki za svaki vremenski segment. Dimenzionalnost rezultirajućeg vektora značajki može se izračunati množenjem broja univarijatnih značajki (15) s brojem odabranih signala (7), te pridodavanjem umnoška broja bivarijantnih značajki (4) i broja odabranih bivarijantnih kombinacija signala (5), što ukupno iznosi $15 \times 7 + 4 \times 5 = 125$ za svaki vremenski interval. U izlaznoj datoteci normalan period označen je s „0“, a napadaj s „1“. Za izlučivanje značajki korišteni su zapisi 12 različitih pacijenata (chb01, chb02, chb03, chb05, chb08, chb11, chb12, chb13, chb14, chb17, chb18,

chb19). Ukupno je generirano 1406 vektora značajki, 707 za normalne i 699 za epileptičke zapise.

5.3.2. Analiza podataka

Točnost klasifikacije procijenjena je korištenjem sustava za dubinsku analizu podataka WEKA, pomoću dva poznata algoritma strojnog učenja: C4.5 i random forest.

5.3.2.1. C4.5

C4.5 je algoritam kojeg je razvio Ross Quinlan, a koristi se za generiranje stabla odlučivanja. Stabla odlučivanja mogu se koristiti za klasifikaciju određenog slučaja, počevši od korijena stabla, kretanjem kroz njega dok se ne naiđe na list, koji određuje predviđeni razred. Na svakoj točki grananja C4.5 odabire atribut koji najefikasnije dijeli skup uzoraka u podskupove koristeći koncept informacijske entropije, te se dalje obrađuje podstablo koje odgovara ishodu ovog testa. Skup podataka za treniranje iz kojeg se formira stablo odlučivanja je skup $S = s_1, s_2, \dots$ već klasificiranih uzoraka. Svaki se uzorak s_i sastoji od p -dimenzionalnog vektora $(x_{1,i}, x_{2,i}, \dots, x_{p,i})$, gdje x_j predstavlja atribut ili značajke uzorka, te razreda u koji s_i upada. U ovom radu korišten je J48, što je implementacija C4.5 algoritma u sustavu WEKA. Pritom su korištene zadane postavke [5].

5.3.2.2. Random forest

Random forest (RF) razvili su Leo Breiman i Adele Cutler [44]. Sadrži više stabla odlučivanja, od kojih svako klasificira ulazni vektor te na taj način „glasuje“ za rezultirajući razred, a na kraju se odabire klasifikacija koja ima najviše glasova. Detaljnije, za svako pojedino stablo u svakom se čvoru nasumično odabire podskup od m atributa, te se odabire atribut koji najefikasnije dijeli promatrani čvor u tom stablu. Za razliku od C4.5, svako stablo u RF se generira u potpunosti, bez rezanja. Prilikom izgradnje stabala stvara internu nepristranu procjenu pogreške klasifikacije (tzv. *out-of-bag (oob) error*), zbog čega ne zahtijeva odvojeni skup za testiranje.

5.3.2.3. Provedba analize pomoću WEKA alata i rezultati

U početnom (*Weka GUI Chooser*) prozoru, odabrana je opcija *Explorer*, koja predstavlja grafičko sučelje koje omogućuje analizu podataka. Grafičko sučelje *Explorer* sadrži više

izbornika: *Preprocess, Classify, Cluster, Associate, Select attributes, Visualize*. Najprije je u izborniku *Preprocess* učitana unaprijed pripremljena datoteka u CSV formatu, koja je izlaz EEGFramea. Prvi redak datoteke čine atributi odvojeni zarezima, pri čemu je posljednji *Class*. U svim ostalim redovima nalaze se podaci, također odvojeni zarezima, pri čemu svaki redak sadrži isti broj stupaca u istom poretku.

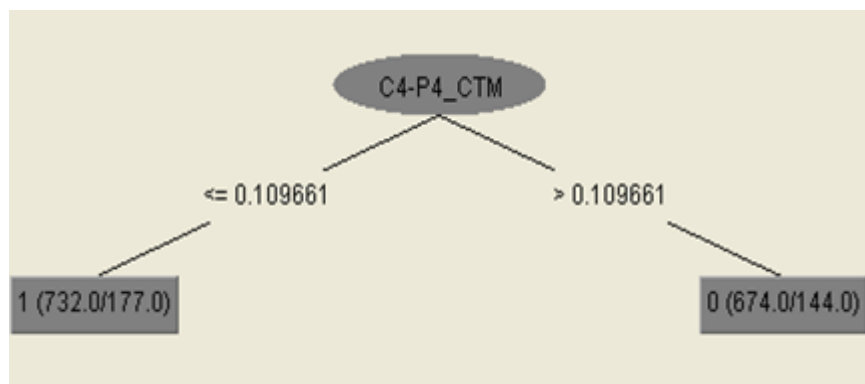
U izborniku *Classify* potrebno je odabrati klasifikator, *Test options*, koji predstavlja evaluaciju klasifikatora, i ciljni atribut (*Class*). Nakon pokretanja klasifikacije u *Classify output* prozoru prikazani su rezultati predviđanja. Klasifikator je evaluiran uz pomoć unakrsne validacije (engl. *cross-validation, CV*), pri čemu je cijeli skup podataka virtualno podijeljen na X podskupova podataka jednake duljine. Svaki podskup podataka je podijeljen na skup podataka za treniranje i skup podataka za testiranje. Korišten je zadani X (tzv. *folds*), koji iznosi 10.

Najprije je odabrano stablo odluke J48 iz paketa *Trees* uz zadane parametre i zabilježena je točnost klasifikacije za sve odabrane značajke, kao i za različite kombinacije. Nakon više provedenih pokusa, izabrani su značajniji rezultati i prikazani u tablici II. Odabirom različitih kombinacija značajki točnost klasifikacije se mijenja. Primjerice, uklanjanjem značajke LZ odmah na početku, točnost klasifikacije se povećala. U nekim drugim kombinacijama, uklanjanje značajke LZ uzrokovalo je smanjenje točnosti klasifikacije. Analizirane su i različite kombinacije signala na kojima su se izlučivale značajke. Većinom su na svim signalima odabrane iste značajke, no za neke se slučajeve diskriminacija pokazala boljom za različite značajke na različitim signalima. Takav je, primjerice, predzadnji redak u tablici, gdje je izabran LZ za signal P3-O1, a CTM za C4-P4.

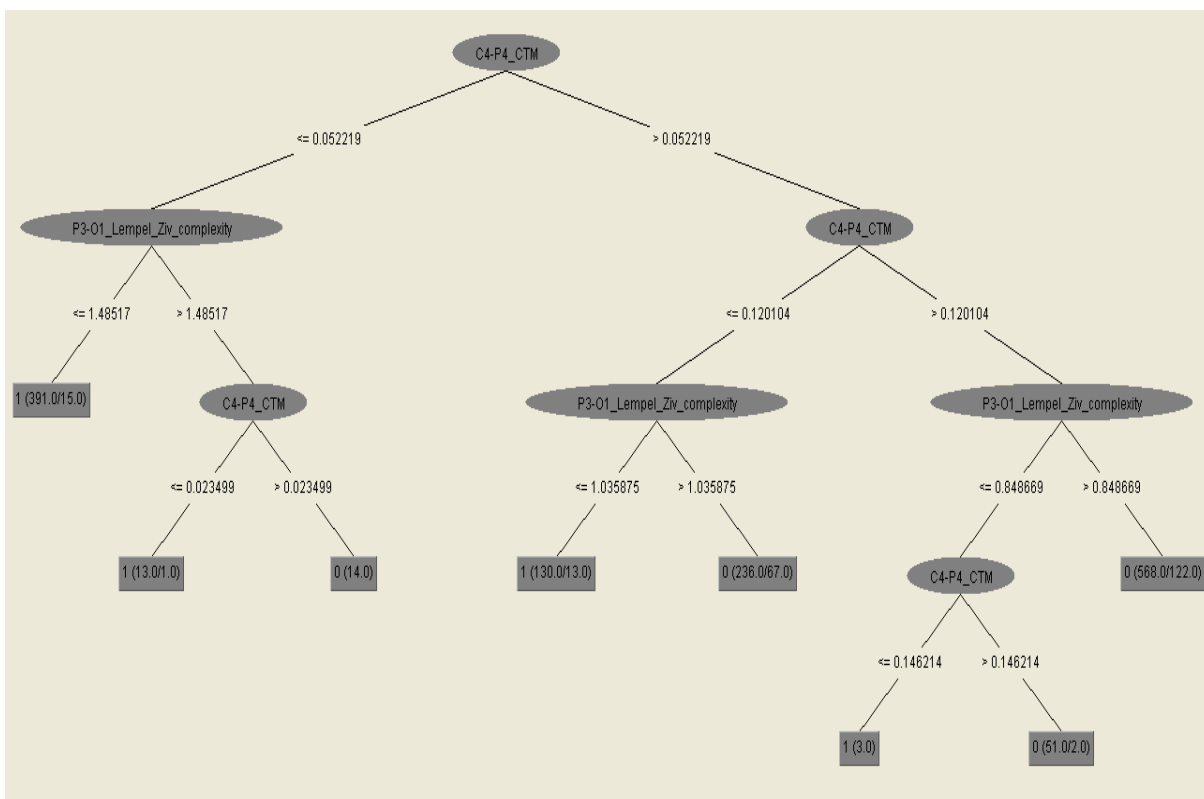
TABLICA II. REZULTATI KLASIFIKACIJE ZA J48

Kombinacije značajki	Analizirani signali	Točnost klasifikacije
Mean, Hurst, D_2 , L_{max} , SampEn(0.1, 0.15, 0.2, 0.25), NLP(1, 2, 3, 4, 5), LZ, CTM, CRP rate, CRP DET, CRP Lam, D_m	F3-C3, F4-C4, P3-O1, P4-O2, T8-P8, C3-P3, C4-P4, F3-C3_C3-P3, F7-T7_T7-P7, F4-C4_C4-P4, F8-T8_T8-P8, FP1-F3_F3-C3	88.9758 %
Mean, Hurst, D_2 , L_{max} , SampEn(0.1, 0.15, 0.2, 0.25), NLP(1, 2, 3, 4, 5), CTM, CRP rate, CRP DET, CRP Lam, D_m	F3-C3, F4-C4, P3-O1, P4-O2, T8-P8, C3-P3, C4-P4, F3-C3_C3-P3, F7-T7_T7-P7, F4-C4_C4-P4, F8-T8_T8-P8, FP1-F3_F3-C3	90.2560 %
Mean, Hurst, D_2 , L_{max} , SampEn(0.1, 0.15, 0.2, 0.25), NLP(1, 2, 3, 4, 5), LZ, CRP rate, CRP DET, CRP Lam, D_m	F3-C3, F4-C4, P3-O1, P4-O2, T8-P8, C3-P3, C4-P4, F3-C3_C3-P3, F7-T7_T7-P7, F4-C4_C4-P4, F8-T8_T8-P8, FP1-F3_F3-C3	78.0228 %
LZ, CTM	F3-C3, F4-C4, P3-O1, P4-O2, T8-P8, C3-P3, C4-P4	89.4026 %
LZ, CTM	F3-C3, P3-O1, P4-O2, T8-P8, C4-P4	90.3272 %
LZ, CTM	P3-O1 (LZ), C4-P4 (CTM)	82.7169 %
CTM	C4-P4	75.7468 %

Zbog veličine grafa, za većinu slučajeva nije bilo moguće vizualizirati stablo odlučivanja. Primjer stabla za 1 atribut (CTM na signalu C4-P4, naveden u posljednjem retku tablice), prikazan je na slici 23, dok je na slici 24 prikazano stablo za attribute LZ na signalu P3-O1 i CTM na signalu C4-P4.



Slika 23. Generirano stablo odlučivanja za klasifikaciju na temelju značajke CTM izračunate na signalu C4-P4



Slika 24. Generirano stablo odlučivanja za LZ na signalu P3-O1 i CTM na signalu C4-P4

Nakon toga je provedena analiza pomoću postupka random forest, koji je također u paketu *Trees*, uz broj stabala 50. Testiranje je najprije izvršeno za kombinaciju svih inicijalno odabranih značajki, a potom na isti način za skup značajki bez CTM-a, za kojeg se pomoću J48 pokazalo da ima posebno velik utjecaj na točnost klasifikacije.

Kako random forest sadrži faktor slučajnosti (zbog odabira m slučajnih atributa u čvoru), program je pokrenut 5 puta za različite vrijednosti čvrste točke (engl. *seed*), izračunata je srednja vrijednost i odstupanje, a rezultati su prikazani za oob pogrešku i točnost klasifikacije pomoću unakrsne validacije (CV) u tablici III.

TABLICA III. REZULTATI KLASIFIKACIJE ZA RANDOM FOREST – SVE ZNAČAJKE

	Seed	Oob	Točnost CV (%)
1.	1	0.1629	91.4651
2.	2	0.1679	92.5320
3.	5	0.1579	92.1053
4.	10	0.1650	92.3186
5.	100	0.1515	91.8208
Srednja vrijednost		0.16104	92.04836
Standardna devijacija		0.00646	0.41897

Rezultati testiranja za sve značajke bez CTM-a prikazani su u tablici IV. Vidljivo je da su lošiji nego za slučaj kad se koristi CTM, što je i očekivano s obzirom na ranije rezultate.

TABLICA IV. REZULTATI KLASIFIKACIJE ZA RANDOM FOREST – CTM ISKLJUČEN

	Seed	Oob	Točnost CV (%)
1.	1	0.2482	85.4908
2.	2	0.2198	84.9218
3.	5	0.2312	84.8506
4.	10	0.2319	86.0597
5.	100	0.2404	84.2105
Srednja vrijednost		0.23430	85.10668
Standardna devijacija		0.01068	0.69977

5.3.3. Analiza rezultata i mogućnosti poboljšanja

Uz više provedenih pokusa s različitim kombinacijama značajki, rezultati su bili vrlo slični, osim u slučaju uklanjanja značajke CTM, kada je točnost klasifikacije značajno pala, te se CTM u svim kombinacijama pokazala vrlo značajnom. Posebno je zanimljiv rezultat u tablici II prema kojem je točnost klasifikacije za skup svih značajki bez CTM-a i sve izabrane signale 78.02 %, dok je točnost klasifikacije za CTM i signal C4-P4 čak 75.75 %. U većini je eksperimenata uklanjanjem značajki faznog prostora točnost klasifikacije rasla, a mogući uzrok su neprikladno odabrani parametri faznog prostora.

Time je pokazano da se pomoću EEGFramea uspješno mogu izlučiti različite kombinacije mnogobrojnih univarijatnih i multivarijatnih značajki, u datoteku formata pogodnog za analizu pomoću nekog od alata namijenjenih dubinskom pretraživanju podataka.

Međutim, postoji veliki broj parametara koji se moraju uzeti u obzir ovisno o prirodi signala koji se analiziraju. Za početak, moguće je izlučiti značajke na svim signalima koji postoje u određenom zapisu ili samo na određenim kombinacijama, pri čemu je efikasnije odabrati signale ovisno o području mozga u kojem se promatrano stanje najintenzivnije manifestira. Razina uključenosti pojedinih kanala u promatranu aktivnost može se jako razlikovati među pacijentima, pa i to predstavlja problem i ne daje nužno očekivane rezultate [43]. Količina podataka potrebna za prikladan opis pojedinog razreda drastično se povećava s dimenzionalnosti vektora značajki, zbog čega će velik broj analiziranih signala u kombinaciji s velikim brojem odabranih značajki znatno otežati i produžiti postupak pripremanja datoteke za daljnju obradu, jer će rezultati klasifikacije uz premalo podataka vjerojatno biti loši. Prikladan odabir značajki posebno je važan, jer neke mnogo više doprinose kvaliteti dobivenih rezultata, te je nužno razlučiti njihov utjecaj. Potrebna je iscrpna analiza, poznavanje prirode stanja koja se analiziraju, algoritama koji se koriste i utjecaja vrijednosti parametara koji se pritom odabiru, te velik broj eksperimenata da bi se razlučila najbolja kombinacija značajki.

Primjena nelinearnih postupaka pretpostavlja postojanje determinističkog kaosa, zbog čega je potrebno najprije analizirati zapise korištenjem zamjenskih podataka, kako bi se odbacila mogućnost pogrešnih rezultata, odnosno pogrešne indikacije prisustva nelinearne dinamike u potpuno slučajnim linearnim sustavima. S druge strane, istraživači često nemaju za cilj određivati postoji li kaos ili ne, te primjenjuju nelinearne značajke neovisno o tome postoji li nelinearnost ili je ona vrlo slaba.

Također, ovisno o odabranim značajkama nužno je podesiti duljinu pomičnog prozora, tako da postoji kompromis između zahtijevane statističke preciznosti za računanje značajke i približne stacionarnosti unutar duljine prozora.

Za sve značajke koje zahtijevaju određene parametre nužno je prikladno ih odrediti, bilo eksperimentalno, bilo na temelju otprije poznatih rezultata. Tako se, primjerice, dimenzija faznog prostora i vremenski odmak prilikom rekonstrukcije faznog prostora za određeni vremenski niz uglavnom određuju iterativnim pokušajima, prvim prolaskom autokorelacijske funkcije kroz nulu i raznim drugim postupcima.

6. Zaključak

Zbog složenosti vremenskih nizova koji se promatraju, analiza EEG signala vrlo je zahtjevna. Veliki broj razvijenih postupaka i značajki, posebno nelinearnih, unaprijedio je razumijevanje osnovnih stanja mozga i poremećaja, kao i izgradnju opisnih i prediktivnih modela za njih. Međutim, postojeći alati koji omogućavaju analizu EEG-a uglavnom podržavaju drugačije postupke i mogućnosti analize, što dodatno otežava taj proces i uzrokuje otežanu usporedbu rezultata istraživača. Očito postoji potreba za programskim produktom koji bi imao mogućnost vizualizacije EEG signala i njihove analize, te pohrane vektora izlučenih značajki u datoteku u formatu pogodnom za otkrivanje znanja, sve na jednom mjestu, što razvijeni radni okvir upravo omogućava.

Najveća prednost EEGFramea je veliki broj podržanih linearnih i nelinearnih postupaka, te vrlo jednostavno generiranje kombinacija mnogo značajki. Time je omogućena opsežna usporedba različitih značajki i pristupa za analizu pojedinih stanja.

Kako je *open-source* i omogućava eksperimentiranje s novim modelima poremećaja, moguće ga je koristiti u medicinske, edukacijske i istraživačke svrhe. Može se upotrijebiti kao samostalan programski produkt ili kao dio većeg sustava, integriranjem samo paketa koji sadrže značajke.

Planira se usavršavanje i prilagodba EEGFramea, dodavanje novih funkcionalnosti, značajki i postupaka, a posebno nelinearnih (npr. fazna sinkronizacija, Hjorth pokretljivost i složenost (engl. *Hjorth mobility and complexity*), SVD entropija, Fisherova informacija itd.).

Lea Suć

7. Literatura

1. Lehnertz, K. „Epilepsy and nonlinear dynamics“, J Biol Phys. 2008;34(3-4): 253–266
2. Mormann, F., Andrzejak, R., Elger, C.E., Lehnertz, K. „Prediction: the long and winding road“, Brain 2007;130(2): 314-333
3. Elger, C.E., Widman, G., Andrzejak, R., Arnhold, J., David, P., Lehnertz, K. „Nonlinear EEG analysis and its potential role in epileptology“, Epilepsia 2000;41 Suppl 3:S34-38
4. Stam, C.J. „Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field“, Clinical Neurophysiology 2005;116:2266-2301
5. Witten, I.H., Frank, E., Hall, M.A. „Data Mining: Practical Machine Learning Tools and Techniques“, 3rd ed., Morgan Kaufmann, 2011.
6. Šantić, A. „Biomedicinska elektronika“, Školska knjiga, 1995.
7. Malmivuo, J., Plonsey, R. „Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields“, Oxford Univ. Press, New York, 1995.
8. Berger, H. „Über das Elektrenkephalogramm des Menschen“, Archiv für Psychiatrie und Nervenkrankheiten, Vol. 87 (1929), pp. 527-570
9. Kemp, B., Värri, A., Rosa, A.C., Nielsen, K.D., Gade, J. „A simple format for exchange of digitized polygraphic recordings“, Electroencephalography and Clinical Neurophysiology, 82 (1992): 391-393.
10. Kemp, B., Olivan, J. „European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data“, Clinical Neurophysiology, 114 (2003): 1755-1761
11. Bravi, A., Longtin, A., JE Seely, A. „Review and classification of variability analysis techniques with clinical applications“, BioMedical Engineering OnLine 2011;10:90
12. Takens, F. „Detecting strange attractors in turbulence“, D. A. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, 1981;898: 366–381
13. Rosenstein MT, Collins JJ, De Luca CJ. „Reconstruction expansion as a geometry-based framework for choosing proper delay times“, Physica D 1994;73:82-98
14. Cellucci CJ, Albano AM, Rapp PE. „Comparative study of embedding methods“, Phys. Rev. E 2003;67:066210

15. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. „Heart Rate Variability“, *Circulation*. 1996;93:1043-1065
16. JE Seely, A., T Macklem, P. „Complex systems and the technology of variability analysis“, *Critical Care* 2004;8:R367-R384
17. Wang, X.-W., Nie, D., Lu, B.-L. “EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines”, *Lecture Notes in Computer Science*, vol. 7062, pp. 734–743, 2011.
18. Teich, M. C., Lowen, S. B., Jost, B. M., Vibe-Rheymer, K., Heneghan, C., “Heart-Rate Variability: Measures and Models”, in *Dynamic Analysis and Modeling*, ser. *Nonlinear Biomedical Signal Processing*, M. Akay, Ed. New York: IEEE Press, 2001, vol. II, ch. 6, pp. 159–213.
19. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, Q., et al., “The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis”, *Proc. of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
20. Rosenstein, M.T., Collins, J.J., De Luca, C.J. „A practical method for calculating largest Lyapunov exponents from small data sets“, *Physica D* 1993;65:117-134
21. Grassberger P, Procaccia I. „Characterization of strange attractors“, *Phys Rev Lett* 1983;50:346-9.
22. G. Sugihara and R. M. May, “Non-linear forecasting as a way of distinguishing chaos from measurement error in time series”, *Nature*, vol. 344, pp. 734–740, 1990.
23. O. Faust, R. U. Acharya, S. M. Krishnan, L. C. Min, “Analysis of cardiac signals using spatial filling index and time-frequency domain”, *BioMed. Eng. OnLine*, vol. 3, p. 30, Sep. 2004.
24. Lin, C.-W. , Wang, J.-S., Chung, P.-C., “Mining Physiological Conditions from Heart Rate Variability Analysis,” *IEEE Computat. Intell. Mag.*, vol. 5, no. 1, pp. 50–58, Feb. 2010.
25. Marwan, N., Romano, M. C., Thiel, M., Kurths, J. „Recurrence Plots for the Analysis of Complex Systems“, *Physics Reports*, 438(5-6), 2007, 237-329

26. Cohen, M.E., Hudson, D.L., Deedwania, P. C. "Applying continuous chaotic modeling to cardiac signal analysis", IEEE Eng. Med. Biol. Mag., vol. 15, no. 5, pp. 97–102, Sep./Oct. 1996.
27. Higuchi, T. "Approach to an irregular time series on the basis of the fractal theory", Physica D, vol. 31, issue 2, pp. 277--283, June 1988.
28. Peng, C.-K., Havlin, S., Stanley, H.E., Goldberger, A.L. "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," Chaos, Solitons, & Fractals, vol. 5, no. 1, pp. 82–87, Jan. 1995.
29. Pincus, S.M., Goldberger, A.L., "Physiological time-series analysis: what does regularity quantify?", Am. J. Physiol., vol. 266, no. 4, (Heart Circ. Physiol., vol. 35), pp. H1643–H1656, Apr. 1994.
30. Richman, J.S., Moorman, J.R., "Physiological time-series analysis using approximate entropy and sample entropy", Am. J. Physiol. (Heart Circ. Physiol.), vol. 278, no. 6, pp. 2039–2049, June 2000.
31. Waheed, K., Salam, F.M., "A Data-Derived Quadratic Independence Measure for Adaptive Blind Source Recovery in Practical Applications," in Proc. 45th IEEE Int. Midwest Symposium on Circuits and Systems, pp. 473–476, 2002.
32. Zhang, X.-S., Zhu, Y.-S., Thakor, N.V., Wang, Z.-Z., "Detecting Ventricular tachycardia and fibrillation by complexity measure," IEEE Trans. Biomed. Eng., vol. 46, no. 5, pp. 548–555, May 1999.
33. Lempel, A., Ziv, J. „On the complexity of finite sequences“, IEEE Trans. Inform. Theory 1976;22:75–81
34. Cohen, M.E., Hudson, D.L., „Inclusion of ECG and EEG. Analysis in neural network models“, Proceedings of International Conference of the 23rd Annual IEEE EMBS October 25-28, 2001, Istanbul Turkey, pp. 1621-1624.
35. Stam, C.J., van Woerkom, T.C.A.M., Pritchard, W.S. „Use of non-linear EEG measures to characterize EEG changes during mental activity“, Electroencephalography and clinical Neurophysiology 99 (1996) 214-224

36. Stam, C.J., van Dijk, B.W. „Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets“, *Physica D* 163 (2002) 236–251
37. Suć, L. „Oblikovanje i implementacija programskog produkta za prikaz elektroencefalografskih signala“, Završni rad, Zagreb 2011.
38. Jović, A., Bogunović, N. „HRVFrame: Java-Based Framework for Feature Extraction from Cardiac Rhythm“, *Lecture Notes in Artificial Intelligence* 6747:96-100, 2011.
39. „PhysioBank“, <http://www.physionet.org/physiobank/>, 28.06.2013.
40. Iasemidis, LD., Sackellares, JCh., Zaveri, HP., Williams, WJ. „Phasespace topography and the Lyapunov exponent of electrocorticograms in partial seizures“, *Brain topogr* 1990;2:187–201
41. Lopes da Silva, FH., Blanes, W., Kalitzin, SN., Parra, J., Suffczynski, P., Velis, DN. „Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity“. *Epilepsia* 2003;44:72–83
42. Mormann, F., Kreuz, Th., Rieke, Ch., Andrzejak, RG., Kraskov, A., David, P., Elger, ChE., Lehnertz, K. „On the predictability of epileptic seizures“, *Clin Neurophysiol* 2005;116:569–97
43. Shoeb, A., Guttag, J. „Application of Machine Learning to Epileptic Seizure Detection“, *Proceedings of the Twenty-seventh International Conference on Machine Learning*, 2010.
44. „Random Forests“, Leo Breiman and Adele Cutler,
http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#intro,
28.06.2013.

Razvrstavanje elektroencefalografskih signala temeljeno na izlučivanju značajki

Sažetak

Razvijen je radni okvir za vizualizaciju elektroencefalografskih (EEG) signala i izlučivanje značajki, nazvan EEGFrame. Implementiran je u Javi i podržava veliki broj linearnih i nelinearnih postupaka opisanih u literaturi. Izlučeni vektori značajki pohranjeni su u datoteku u CSV formatu, pogodnom za dubinsku analizu podataka. Pomoću EEGFramea analizirane su datoteke u standardnom EDF formatu, koje sadrže epileptičke zapise 12 mladih pacijenata i djece. Izabrana je kombinacija linearnih i nelinearnih univarijatnih i bivarijatnih značajki. Otkrivanje znanja provedeno je korištenjem *open-source* platforme WEKA. Cilj je bio detekcija napadaja provedena kao binarna klasifikacija (EEG napadaja ili normalan EEG). Točnost klasifikacije procijenjena je pomoću dva poznata algoritma strojnog učenja: C4.5 i random forest. Rezultati su za odabranu kombinaciju značajki bili obećavajući (89 % s C4.5 i oko 92 % s random forest). Pokazalo se da značajka CTM ima najveći doprinos točnosti izgrađenih modela.

Ključne riječi

elektroencefalografija, izlučivanje značajki, biomedicinski inženjering, dubinska analiza, epilepsija

Feature extraction based classification of electroencephalographic signals

Abstract

Joint framework for electroencephalographic (EEG) signals visualization and feature extraction, called EEGFrame, was developed. EEGFrame is implemented in Java and supports numerous linear and nonlinear measures described in literature. Extracted feature vectors are stored to a file in CSV format, suitable for data mining. Files in standard EDF format containing epileptic data from 12 pediatric patients were analysed using EEGFrame. Combination of linear and nonlinear univariate and bivariate features was selected. Knowledge discovery was performed using *open-source* platform called WEKA. The goal was seizure detection carried out as a binary classification (seizure or non-seizure EEG). Classification accuracy is evaluated by two well-known machine learning algorithms: C4.5 and random forest. Results for the proposed feature combination were promising (89 % with C4.5 and about 92 % with random forest). It was showed that feature CTM has the largest contribution to the accuracy of constructed models.

Keywords

electroencephalography, feature extraction, biomedical engineering, data mining, epilepsy