

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 992

**Klasifikacija afektivnih poremećaja  
korištenjem fraktalne dimenzije  
kao značajke signala EEG-a**

Zvonimir Žunić

Zagreb, lipanj 2023.

*Ovom prilikom želim izraziti svoju iskrenu zahvalnost prof. dr. sc. Mariu Cifreku i mag. ing. Edi Jovičić na njihovom stručnom vođenju, podršci i mentorstvu tijekom izrade ovog rada.*

# SADRŽAJ

<b>1. Elektroencefalografija</b>	<b>1</b>
1.1. Osnove elektroencefalografije . . . . .	1
<b>2. Strojno učenje</b>	<b>3</b>
2.1. Unakrsna validacija . . . . .	3
2.2. Mjere vrednovanja modela strojnog učenja . . . . .	4
2.3. Stablo odluke . . . . .	5
2.4. K-najbližih susjeda (KNN) . . . . .	6
2.5. Hiperparametrizacija modela klasifikacije . . . . .	7
<b>3. Fraktalna dimenzija</b>	<b>9</b>
3.1. Izračun fraktalne dimenzije . . . . .	9
3.2. Tipovi fraktalne dimenzije . . . . .	10
3.2.1. Higuchijeva fraktalna dimenzija . . . . .	10
3.2.2. Katzova fraktalna dimenzija . . . . .	11
3.2.3. Sevcikova fraktalna dimenzija . . . . .	11
3.2.4. Petrosianova fraktalna dimenzija . . . . .	12
<b>4. Skup podataka EEG-a</b>	<b>13</b>
<b>5. Predobrada signala EEG-a</b>	<b>14</b>
<b>6. Izdvajanje značajki signala korištenjem fraktalnih dimenzija</b>	<b>15</b>
<b>7. Klasifikacija afektivnih poremećaja na temelju izdvojenih značajki</b>	<b>16</b>
7.1. Stablo odluke . . . . .	16
7.2. K-najbližih susjeda (KNN) . . . . .	17
7.3. Usporedba modela . . . . .	18
<b>Literatura</b>	<b>20</b>

# UVOD

Afektivni poremećaji uključuju širok spektar mentalnih bolesti poput depresije i bipolarnog poremećaja. Ovi poremećaji, koji primarno utječu na raspoloženje i emocije pojedinca, mogu značajno utjecati na kvalitetu života i svakodnevno funkcioniranje. Unatoč napretku u psihijatriji, dijagnoza afektivnih poremećaja još uvijek je složen proces koji često zahtijeva dugotrajne kliničke procjene. U ovom radu, predstavljamo pristup klasifikaciji afektivnih poremećaja koristeći elektroencefalografiju (EEG) i fraktalnu dimenziju kao značajku EEG signala.

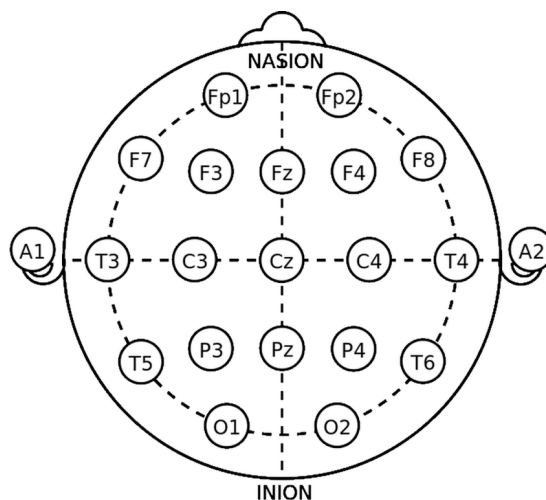
Strojno učenje je područje umjetne inteligencije koje se bavi razvojem algoritama koji poboljšavaju svoju izvedbu kroz iskustvo. Koristimo nadzirane tehnike učenja, posebno stabla odluke (engl. Decision Tree) i K-najbližih susjeda (engl. K-nearest neighbours, KNN), za klasifikaciju afektivnih poremećaja. Ove tehnike omogućuju nam da izvučemo složene obrasce iz EEG podataka i koristimo ih za predviđanje afektivnih poremećaja.

Ovaj rad pruža detaljan opis našeg pristupa. Počinjemo s osnovama elektroencefalografije, metode koja se koristi za snimanje električne aktivnosti mozga, i strojnog učenja. Zatim se fokusiramo na koncept fraktalne dimenzije, matematičkog svojstva koje pruža kvantifikaciju složenosti signala, i kako se ona izračunava. Diskutiramo o procesu predobrade EEG signala, koji je neophodan za uklanjanje smetnji, i izdvajanje značajki koristeći fraktalne dimenzije. Na kraju, opisujemo klasifikaciju afektivnih poremećaja koristeći izdvojene značajke. Svako poglavlje pruža dublje razumijevanje svake faze našeg istraživanja.

# 1. Elektroencefalografija

## 1.1. Osnove elektroencefalografije

Elektroencefalografija (EEG) predstavlja metodu kojom se bilježe električni signali koji nastaju kao rezultat moždane aktivnosti. Elektrode se postavljaju na površinu vlasišta kako bi detektirale električne impulse koji proizlaze iz interakcije neurona. EEG se široko koristi u medicinskim i znanstvenim istraživanjima za proučavanje funkcija mozga, dijagnozu neuroloških poremećaja i razumijevanje kognitivnih procesa. Prilikom provođenja EEG snimanja, elektrode se postavljaju na vlasište prema međunarodno prihvaćenom rasporedu elektroda poznatom kao 10-20 sustav. Ovaj sustav omogućuje precizno pozicioniranje elektroda kako bi se osigurala pouzdana i usporediva mjerenja između različitih subjekata i istraživanja [1]. Moždani signali obuhvaćaju različite frekvencijske komponente koje se odnose na različite razine moždane aktivnosti. Ovi ritmovi klasificiraju se kao delta, teta, alfa, beta i gamma valovi, a svaki od njih povezan je s određenim funkcionalnim stanjima mozga.



**Slika 1.1:** Sustav montaže 10-20 [2]

Gamma moždani valovi (30 Hz - 100 Hz) moždane valove karakterizira najveća

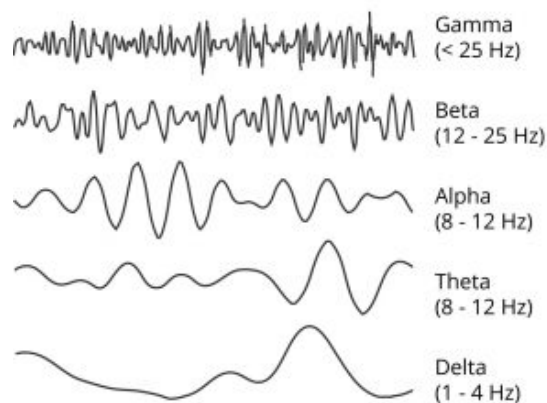
brzina i najmanja amplituda. Oni su povezani s višim kognitivnim funkcijama, kao što su percepcija, rješavanje problema i svijest. Gamma valovi sudjeluju u učenju, pamćenju i obradi informacija.

Beta moždani valovi (12 Hz - 30 Hz) moždani valovi povezani su s budnim i fokusiranim mentalnim stanjem. Prevladavaju tijekom zadataka koji zahtijevaju koncentraciju, kao što su rješavanje problema i donošenje odluka. Prekomjerna proizvodnja beta valova može biti povezana s anksioznošću i stresom.

Alfa moždani valovi (8 Hz - 12 Hz) moždani valovi su sporiji od beta valova i povezani su s opuštenim i smirenim mentalnim stanjem. Oni su prisutni tijekom aktivnosti poput sanjarenja, lagane meditacije i kreativne vizualizacije. Alfa valovi pomažu nam u tranziciji između svjesnog i podsvjesnog uma, te igraju ulogu u poboljšanju cjelokupne mentalne koordinacije, smirenosti i budnosti.

Teta moždani valovi (4 Hz - 8 Hz) moždani valovi su još sporiji od alfa valova i povezani su s dubokom opuštenošću, meditacijom i ranim stadijima sna. Također su povezani s povećanom kreativnošću, pamćenjem i učenjem. Teta valovi mogu biti prisutni tijekom aktivnosti poput sanjarenja, duboke meditacije i REM sna.

Delta moždani valovi (0,5 Hz - 4 Hz) moždani valovi su najsporiji i imaju najveću amplitudu. Povezani su s dubokim snom i oslobađanjem hormona rasta. Delta valovi su ključni za procese obnove tijela i cjelokupnog ozdravljenja. Također igraju ulogu u formiranju dugotrajnih sjećanja i regulaciji nesvjesnih tjelesnih procesa [3] [4].



**Slika 1.2:** Moždani valovi [5]

Tradicionalne metode dijagnostike afektivnih poremećaja uglavnom se temelje na kliničkim intervjuima i procjeni simptoma, što može biti subjektivno i nepouzđano. U posljednjih nekoliko desetljeća, istraživači su počeli istraživati potencijalnu primjenu EEG-a kao dodatne metode za dijagnostiku afektivnih poremećaja.

## 2. Strojno učenje

Obično se pod izrazom strojno učenje misli na izmjene u sustavima koji obavljaju zadatke povezane s umjetnom inteligencijom (engl. Artificial Intelligence, AI). Primjerice, takvi zadaci mogu uključivati prepoznavanje, dijagnosticiranje, planiranje, upravljanje robotima, predviđanje i slično. Kada se govori o "izmjenama", one se odnose na moguća poboljšanja postojećih sustava ili kreiranje novih sustava od nule [6]. Postoje dvije vrste strojnog učenja: nadzirano učenje i nenadzirano učenje. Nadzirano učenje je metoda strojnog učenja gdje se sustav trenira na označenim skupovima podataka s poznatim ulaznim i izlaznim varijablama. Cilj nadziranog učenja je predviđanje ishoda za nove podatke. Ova metoda je idealna za zadatke kao što su otkrivanje neželjene pošte, analiza sentimenta, predviđanje vremena i predviđanje cijena. Nadzirano učenje zahtijeva veliki broj označenih skupova podataka za uspješno razumijevanje uzoraka i odnosa unutar njih. Nenadzirano učenje je metoda strojnog učenja gdje se algoritmu daju neklasificirani podaci. Cilj nenadziranog učenja je dobivanje uvida iz velikih količina novih podataka identificiranjem strukture unutar zadanih ulaznih podataka. Nenadzirano učenje ne zahtijeva toliko označenih skupova podataka za učenje kao nadzirano učenje jer se fokusira na identificiranje strukture unutar svojih zadanih ulaznih podataka umjesto pokušaja podudaranja predefiniranog obrasca [7]. Dvije glavne upotrebe nadziranog učenja su klasifikacija i regresija. Ključna razlika između ove dvije metode je u tome što je izlaz klasifikacijskog algoritma klasa, dok je izlaz regresije stvarna vrijednost [8]. U ovom radu koristit ćemo python biblioteku scikit-learn za implementaciju modela strojnog učenja.

### 2.1. Unakrsna validacija

Unakrsna validacija (engl. Cross-Validation) je statistička metoda koja se koristi za procjenu performansi ili točnosti modela strojnog učenja. Uključuje podjelu dostupnih podataka na više skupina ili podskupina, korištenje jedne od tih skupina kao validacijskog seta, te treniranje modela na preostalim skupinama. Unakrsna validacija se koristi

za zaštitu od prenaučivosti, posebno u slučajevima kada je količina podataka ograničena, te za usporedbu različitih modela strojnog učenja i određivanje koliko dobro rade na rješavanju određenog problema.

Popularna tehnika unakrsne validacije je unakrsna validacija sa k preklopa (engl. k-fold Cross-Validation), koja uključuje podjelu podataka na k jednakih dijelova ili skupina, treniranje modela na k-1 skupina, i testiranje na preostaloj skupini. Ovaj proces se ponavlja k puta, pri čemu svaka skupina jednom služi kao testni set, a uzima se prosjek rezultata kako bi se procijenila performansa modela [9] [10].

## 2.2. Mjere vrednovanja modela strojnog učenja

Mjere vrednovanja koje ćemo koristiti u ovom radu su točnost na testnom skupu (engl. accuracy), preciznost (engl. precision), odziv (engl. recall), F1 i matrica konfuzije (engl. confusion matrix). Točnost modela na testnom skupu je udio ispravno klasificiranih primjera u testnom skupu. Preciznost je udio ispravno klasificiranih pozitivnih primjera među svim primjerima koji su klasificirani kao pozitivni. Odziv je udio ispravno klasificiranih pozitivnih primjera među svim stvarnim pozitivnim primjerima. F1 rezultat je harmonijska sredina preciznosti i osjetljivosti.

		Predicted Class	
True Class	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

**Slika 2.1:** Matrica konfuzije [11]

Matrica konfuzije je tablica koja prikazuje broj ispravno i pogrešno klasificiranih primjera za svaku klasu. Dijagonalni elementi matrice predstavljaju broj ispravno klasificiranih primjera za svaku klasu, dok ostali elementi predstavljaju greške klasifikacije.



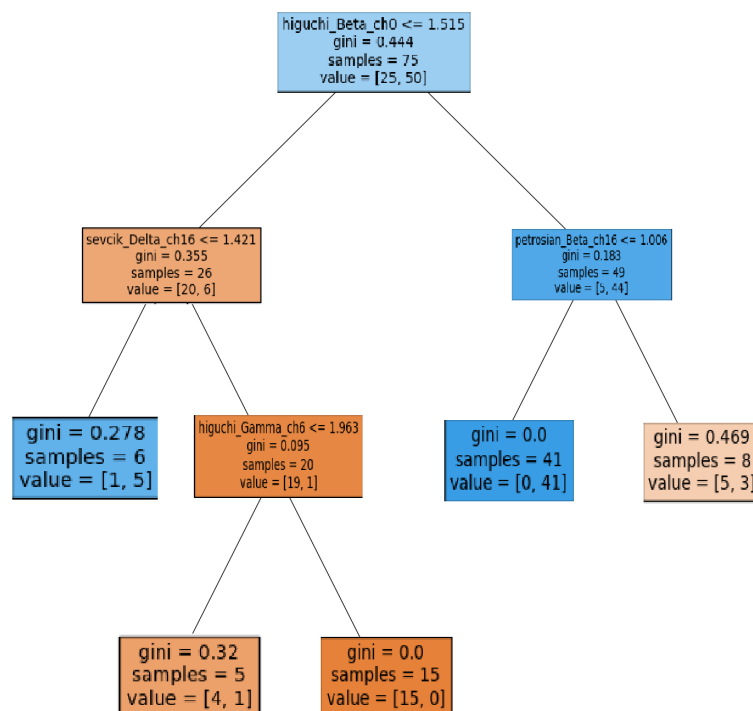
## 2.3. Stablo odluke

Stablo odluke je hijerarhijska struktura koja se koristi u algoritmima strojnog učenja za donošenje odluka prolaskom kroz čvorove. Svaki čvor sadrži atribut ili značajku koja se dalje dijeli na više čvorova kako se krećemo niz stablo. Da bismo odlučili na kojim značajkama ćemo podijeliti stablo, koristimo mjere podjele poput Gini indeksa, dobitka informacija i drugih [12].

Gini indeks, ili Gini nečistoća, je metrika koja se koristi u stablu odluke za odabir značajki za razdvajanje podataka. Gini indeks mjeri vjerojatnost pogrešne klasifikacije kada se slučajno odabere element iz skupa podataka. Drugim riječima, Gini indeks mjeri "nečistoću" skupa podataka - skup podataka je "čist" (Gini indeks je 0) ako svi elementi pripadaju istoj klasi, a "nečist" (Gini indeks je 0,5) ako su elementi ravnomjerno raspoređeni među različitim klasama. Značajka koja daje najmanji Gini indeks (tj. najveću "čistoću") odabire se za razdvajanje. Ovaj postupak se ponavlja rekurzivno za svaki novi podskup podataka dok se stablo ne izgradi do kraja ili dok se ne postigne određeni uvjet zaustavljanja [13].

Graf stabla odluke prikazuje nekoliko važnih informacija. "gini" označava Ginijevu nečistoću koju smo objasnili u prethodnom tekstu. Oznaka "value" predstavlja polje koje pokazuje broj uzoraka svake klase koji su dosegli određeni čvor. Na primjer, u problemu binarne klasifikacije, polje vrijednosti kao što je [25, 50] na čvoru, što znači da od uzoraka koji su dosegli ovaj čvor, 25 pripada jednoj klasi, a 50 drugoj. U terminalnim čvorovima (također poznatim kao listovi), najčešća klasa se koristi kao predviđanje za sve uzorke koji dosegnu taj čvor.

Za implementaciju stabla odluke koristili smo python biblioteku scikit-learn, također koristili smo metodu pretraživanja rešetke (engl. GridSearchCV) za optimizaciju hiperparametara modela, više o tome u sljedećem poglavlju. Nakon što smo pronašli najbolje hiperparametre, koristili smo metodu za odabir značajki (engl. SelectFrom-Model) iz python biblioteke scikit-learn da bismo odabrali najvažnije značajke za klasifikaciju. Ova metoda koristi model stabla odluke da bi odredila važnost svake značajke i odabrala one značajke koje najviše pridonose točnosti modela. Nakon toga, trenirali smo model stabla odluke koristeći odabrane značajke i najbolje hiperparametre. Također smo izračunali točnost modela na skupu za testiranje kako bismo procijenili njegovu generalizacijsku sposobnost.



Slika 2.2: Stablo odluke

## 2.4. K-najbližih susjeda (KNN)

KNN je algoritam strojnog učenja koji se koristi za probleme klasifikacije i regresije u nadziranom učenju. To je jedan od najjednostavnijih algoritama strojnog učenja. Algoritam KNN djeluje tako da pronalazi udaljenosti između čvorova i svih primjera u podacima, odabire određeni broj primjera (K) najbližih čvoru, a zatim glasa za najčešću oznaku (u slučaju klasifikacije) [14].

Osnovna ideja algoritma je da se objekt klasificira na temelju većine klasa njegovih K najbližih susjeda. K je hiperparametar algoritma koji predstavlja broj susjeda koji se uzimaju u obzir prilikom donošenja odluke o klasifikaciji.

Udaljenost između objekata se računa koristeći različite metrike, najčešće euklid-

sku ili manhattansku. Euklidska udaljenost:

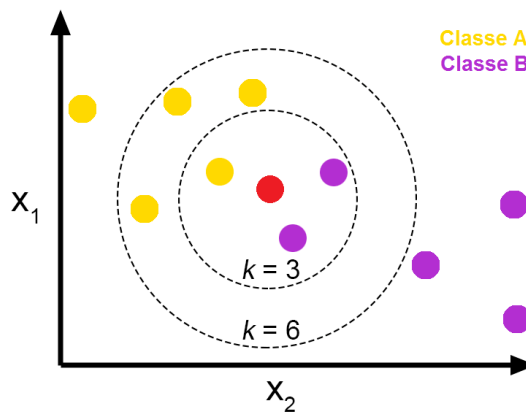
$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (2.1)$$

gdje su  $\mathbf{p}$  i  $\mathbf{q}$  točke u  $n$ -dimenzionalnom prostoru.

Manhattanska udaljenost:

$$d(\mathbf{p}, \mathbf{q}) = |q_1 - p_1| + |q_2 - p_2| + \dots + |q_n - p_n| \quad (2.2)$$

gdje su  $\mathbf{p}$  i  $\mathbf{q}$  točke u  $n$ -dimenzionalnom prostoru.



**Slika 2.3:** Knn model [15]

Skup podataka sa slike 8.2 uključuje ljubičaste kuglice, žute kuglice i kuglicu čija boja je nepoznata. Cilj je odrediti boju nepoznate kuglice na temelju njenih karakteristika, koje su poznate, ali ne i njena oznaka boje. Algoritam KNN mjeri udaljenost (koristeći metriku udaljenosti poput euklidske ili manhattanske) između nepoznate kuglice i svake od poznatih kuglica. Algoritam zatim provjerava boju kuglica koje su imale najkraće udaljenosti i broji broj svake boje. Boja koja se najviše puta pojavljuje uzima se kao ispravna boja za nepoznatu kuglicu. Nepoznata kuglica se zatim klasificira s tom bojom.

Za implementaciju KNN algoritma koristili smo biblioteku scikit-learn, također smo koristili metodu GridSearchCV za optimizaciju hiperparametara. Nakon što smo pronašli najbolje hiperparametre, trenirali smo model i izračunali točnost na test setu.

## 2.5. Hiperparametrizacija modela klasifikacije

Hiperparametrizacija u strojnom učenju je postupak odabira najboljeg skupa hiperparametara za danu arhitekturu modela kako bi se optimizirala njegova izvedba na određenom zadatku. Hiperparametar je parametar čija se vrijednost postavlja prije početka

procesa treniranja i ne može se naučiti iz podataka za treniranje. Umjesto toga, obično ih ručno postavljamo ili se odabiru putem procesa pretraživanja [16].

Hiperparametri koje smo koristili za optimizaciju modela stabla odluke su: `max_depth`, `min_samples_split` i `min_samples_leaf`. `max_depth` je hiperparametar koji određuje maksimalnu dubinu stabla. Dubina stabla je najveći broj uzastopnih veza od korijena do najdaljeg lista. `min_samples_split` je hiperparametar koji određuje minimalni broj uzoraka potrebnih za razdvajanje unutarnjeg čvora. `min_samples_leaf` je hiperparametar koji određuje minimalni broj uzoraka potrebnih da bi čvor bio list.

Hiperparametrizacija za algoritam KNN su: `n_neighbors`, `weights` i `metric`. `n_neighbors` je hiperparametar koji određuje broj najbližih susjeda koji se uzimaju u obzir prilikom donošenja odluke o klasifikaciji. `weights` je hiperparametar koji određuje kako se težine dodjeljuju susjedima prilikom donošenja odluke o klasifikaciji. Može biti `uniform`, što znači da svi susjedi imaju jednaku težinu, ili `distance`, što znači da susjedi koji su bliže imaju veću težinu. `metric` je hiperparametar koji određuje metriku koja se koristi za izračunavanje udaljenosti između uzoraka. Može biti `euclidean`, što znači da se koristi euklidska udaljenost, ili `manhattan`, što znači da se koristi Manhattan udaljenost.

Implementaciju hiperparametara postigli smo metodom pretraživanja rešetke (engl. `GridSearchCV`) iz biblioteke `scikit-learn`. Prvo smo definirali skup hiperparametara za optimizaciju modela. Zatim smo koristili metodu `GridSearchCV` za pronalaženje najbolje kombinacije hiperparametara koji daju najveću točnost modela na skupu za treniranje. Ova metoda provodi iscrpnu pretragu kroz sve moguće kombinacije hiperparametara i odabire onu kombinaciju koja daje najbolje rezultate.

## 3. Fraktalna dimenzija

Fraktalna dimenzija je mjera složenosti fraktala, koji je vrsta matematičkog oblika koji je beskonačno složen i samosličan. Fraktalna dimenzija se može odnositi na bilo koju od dimenzija koje se obično koriste za karakteriziranje fraktala. Viša fraktalna dimenzija može ukazivati na veću složenost signala, dok niža fraktalna dimenzija može ukazivati na manju složenost ili redovitost signala.

### 3.1. Izračun fraktalne dimenzije

Da bismo odredili dimenziju fraktala, određujemo faktor skaliranja  $S$  i broj kopija  $C$  originalnog oblika potrebnih, a zatim koristimo formulu:

$$D = \frac{\log(C)}{\log(S)} \quad (3.1)$$

Faktor skaliranja ovisi o dimenziji oblika. Ako imamo liniju duljine 1 i želimo povećati njenu duljinu za faktor 2, trebat ćemo dvije kopije originalne linije. Ako imamo pravokutnik duljine 2 i visine 1, i želimo povećati njegovu duljinu i širinu za faktor 2, trebat ćemo četiri kopije originalnog pravokutnika. Ako imamo kocku s duljinom stranice 1 i želimo povećati duljinu, širinu i visinu za faktor 2, trebat će nam osam kopija originalne kocke. Navedeni primjeri nas dovode do sljedeće formule kojom povezujemo broj kopija,  $C$  i faktor skaliranja  $S$ .

$$C = S^{\text{dimenzija}} \quad (3.2)$$

Odnosno za jednu dimenziju broj kopija će biti  $S^1$ , za dvije dimenzije  $S^2$  itd [17].

## 3.2. Tipovi fraktalne dimenzije

### 3.2.1. Higuchijeva fraktalna dimenzija

Higuchijeva fraktalna dimenzija (HFD) je metoda za mjerenje fraktalne dimenzije vremenskih nizova podataka. HFD se koristi u raznim primjenama u osnovnom i kliničkom istraživanju neurofiziologije. Koristi se za mjerenje složenosti neuronske aktivnosti i klasificiranje moždanih signala [18]. Matematički, HFD se može izraziti kroz sljedeće korake:

1. Prvo, konstruira se vremenski niz  $X$  iz danog skupa podataka.
2. Zatim se za svaki vremenski interval  $k$  konstruiraju novi vremenski nizovi  $X_k$ , počevši od različitih početnih vremena  $m$ :

$$X_k^m = \{X(m), X(m+k), X(m+2k), \dots, X(m+jk)\}, \quad \text{za } m = 1, 2, \dots, k \quad (3.3)$$

3. Izračunava se duljina svake krivulje  $X_k$ , označena kao  $L_m(k)$ :

$$L_m(k) = \frac{1}{k} \sum_{i=1}^{[(N-m)/k]} |X(m+ik) - X(m+(i-1)k)| \quad (3.4)$$

4. Izračunava se prosječna duljina krivulja za svaki vremenski interval  $k$ , označena kao  $L(k)$ :

$$L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k) \quad (3.5)$$

5. Ako je  $L(k)$  proporcionalno  $k^{-D}$ , tada je krivulja fraktal s dimenzijom  $D$ :

$$L(k) \propto k^{-D} \quad (3.6)$$

### 3.2.2. Katzova fraktalna dimenzija

Katzova fraktalna dimenzija jedna je od metoda koja se koristi za procjenu fraktalne dimenzije. Uvedena je od strane Mosesa Katza 1988. kao mjera nepravilnosti krivulje ili vremenskog niza, kao što je EEG. Katzova fraktalna dimenzija definirana je kao omjer logaritma ukupne duljine krivulje ili broja podataka u vremenskom nizu prema logaritmu maksimalne udaljenosti između bilo koje dvije točke na krivulji ili vremenskom nizu [19]. Matematički, to se može izraziti kao:

$$D = \frac{\log(L)}{\log(d)} \quad (3.7)$$

### 3.2.3. Sevcikova fraktalna dimenzija

Sevcikova fraktalna dimenzija (SFD) je metoda za izračunavanje fraktalne dimenzije valnog oblika. Predložio ju je Carlos Sevcik 1998. godine. SFD algoritam se može koristiti za brzo mjerenje složenosti i nasumičnosti signala. SFD algoritam se koristi u različitim poljima kao što su obrada signala i obrada slika.

SFD algoritam se koristi za procjenu fraktalne dimenzije valnog oblika mjerenjem duljine krivulje valnog oblika. Duljina krivulje izračunava se zbrajanjem udaljenosti između susjednih točaka u valnom obliku. Fraktalna dimenzija se zatim izračunava primjenom Sevcikove formule na duljinu krivulje valnog oblika.

Prvo, valni oblik je podvrgnut dvostrukoj linearnoj transformaciji koja ga preslikava u jedinični kvadrat. Normalizirana apscisa kvadrata je  $x^*$  a normalizirana ordinata je  $y^*$ , oba definirana kao:

$$x^* = \frac{x}{x_{\max}} \quad (3.8)$$

$$y^* = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (3.9)$$

gdje je  $x_{\max}$  maksimalni  $x$ , a  $y_{\min}$  i  $y_{\max}$  su minimalni i maksimalni  $y$ .

Zatim se fraktalna dimenzija valnog oblika ( $\Phi$ ) aproksimira kao  $D$ :

$$\Phi \approx D = 1 + \frac{\ln(L)}{\ln(2 \cdot N')} \quad (3.10)$$

gdje je  $L$  duljina krivulje u jediničnom kvadratu, a  $N' = N - 1$  [20] [21].

### 3.2.4. Petrosianova fraktalna dimenzija

Petrosianova fraktalna dimenzija omogućuje brzi izračun fraktalne dimenzije danog signala pretvaranjem serije u binarnu sekvencu. Binarna sekvencija se formira dodjeljivanjem '1' za svaku razliku između uzastopnih uzoraka u vremenskoj seriji koja premašuje veličinu standardne devijacije, inače se dodjeljuje '0'. Zatim se fraktalna dimenzija izračunava pomoću sljedeće formule:

$$P = \frac{\log_{10}(N)}{\log_{10}(N) + \log_{10}\left(\frac{N}{N+0.4N\sigma}\right)} \quad (3.11)$$

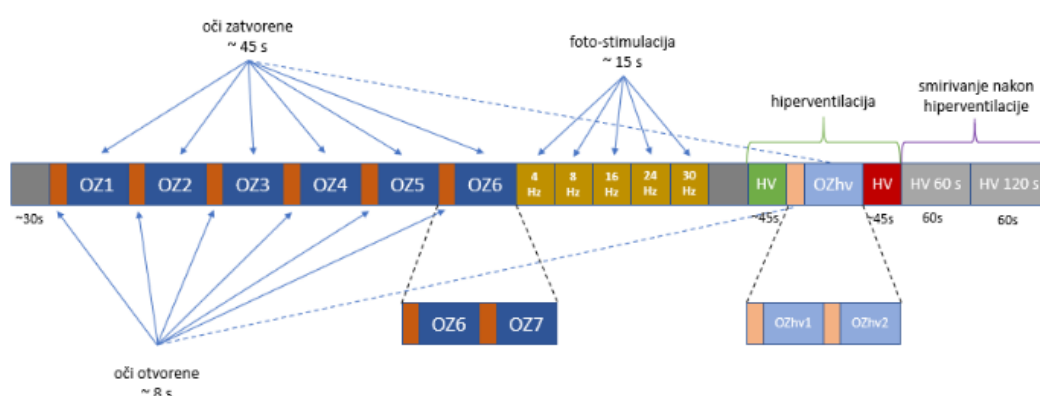
gdje je  $N$  duljina vremenske serije, a  $N\sigma$  je broj promjena znaka u derivatu signala [22].



## 4. Skup podataka EEG-a

Korištena je baza podataka nastala u Klinici za psihijatriju Vrapče. Bazu podataka podijelili smo na set za učenje i set za testiranje. Set za učenje u sebi ima 50 bolesnih ispitanika i 25 zdravih, dok set za testiranje ima 20 bolesnih i 10 zdravih ispitanika. Svakom ispitaniku dodijeljen je jedinstveni identifikator. Dijagnoze ispitanika su F33 (povratni depresivni poremećaj) i F32 (depresija). Dijagnoze F32 i F33 dijele se još i na blage, umjerene i teške.

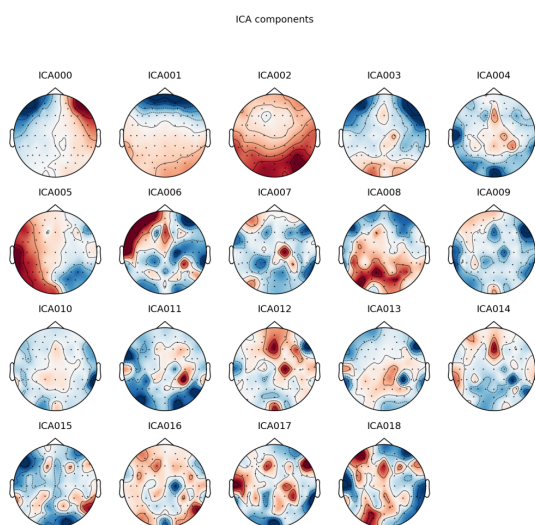
Snimanje EEG-a odvija se po definiranom protokolu. Ispitanik se nalazi u mirnoj i tihoj prostoriji u ležećem položaju. Na glavu mu se postavlja 19 elektroda po unaprijed specificiranom položaju. Protokol snimanja sastoji se od tri dijela: stanje mirovanja, fotostimulacija, hiperventilacija. U stanju mirovanja ispitanik naizmjenice otvara i zatvara oči šest do sedam puta u periodima od 8 i 45 sekundi. Fotostimulacija sastoji se od različitih frekvencija bljeskanja (4 Hz, 8 Hz, 16 Hz, 24 Hz i 30 Hz) od kojih svaka traje 15 sekundi.



Slika 4.1: Tijek snimanje EEG-a

## 5. Predobrada signala EEG-a

Predobrada signala EEG-a potrebna je zbog uklanjanja smetnji. Predobrada signala transformira sirove podatke u nešto što je prikladnije za daljnju analizu podataka. Najčešće smetnje su: gradska mreža, treptanje očiju, pokret pacijenta, pomak elektroda. Prvi korak predobrade signala je filtriranje signala u frekvencijskom rasponu od 0,1 Hz do 40 Hz korištenjem python funkcije `filter` iz MNE biblioteke. Odabrali smo navedeni frekvencijski raspon jer se u tom rasponu nalaze moždani valovi od interesa. Sljedeći korak je primjena nezavisne analize komponenti (engl. Independent Component Analysis, ICA) kako bi se identificirale komponente povezane sa smetnjama poput treptanja očiju. Za ovaj korak se koristi funkcija `ICA` iz MNE biblioteke.



**Slika 5.1:** Primjer ICA-e

Nakon što se izvrši ICA, potrebno je pronaći artefakte u signalu koje je uzrokovalo treptanje očiju. To se radi korištenjem funkcije `find_bads_eog`. Pronađeni artefakti se zatim isključuju iz daljnje obrade korištenjem varijable `exclude`. Nakon ICA-e filtrirali smo signal u pet frekvencijskih raspona, za svaki moždani val jedan frekvencijski raspon. Tako smo iz jednog signala dobili pet dijelova od kojih svaki prikazuje jedan moždani val.

## 6. Izdvajanje značajki signala korištenjem fraktalnih dimenzija

U ovom poglavlju detaljno ćemo opisati postupak izdvajanja značajki iz EEG signala koristeći fraktalne dimenzije. Početni korak uključuje učitavanje i predobradu EEG signala.

EEG signal je snimljen s 19 kanala, što znači da svaki uzorak sadrži 19 različitih mjerenja električne aktivnosti mozga. Svaki tip moždanog vala obrađen je zasebno. Za svaki val, podaci su izdvojeni i analizirani po kanalima. Za svaki od 19 kanala, izračunate su četiri različite fraktalne dimenzije - Higuchijeva, Sevcikova, Katzova i Petrosianova. Ove fraktalne dimenzije služe kao značajke za daljnju analizu.

Nakon izračuna fraktalnih dimenzija, iz imena datoteke izvučen je identifikacijski broj ispitanika. Za svaki kanal svakog moždanog vala, izračunate značajke su spremljene u strukturu podataka, uključujući identifikacijski broj ispitanika, ime datoteke i izračunate fraktalne dimenzije. Ove značajke su zatim dodane u skup podataka za daljnju obradu i analizu.

Kako bi se omogućila klasifikacija, dodali smo informacije o dijagnozi subjekata u skup podataka s izdvojenim značajkama. Informacije o dijagnozi subjekata učitane su iz metapodataka koji su spremljeni u .xlsx formatu. Dodali smo stupac oznake (engl. label) u postojeći skup podataka. Ovaj stupac sadrži vrijednosti 0 ili 1, ovisno o tome je li ispitanik zdrav ili ima afektivni poremećaj. Tako smo pripremili podatke za daljnje korake.

Nakon što smo dodali informacije o dijagnozi i izračunali fraktalne dimenzije za svaki kanal svakog moždanog vala, dobili smo konačne skupove podataka za treniranje i testiranje. Dimenzije konačnog skupa podataka za treniranje su (75, 383), što znači da skup sadrži 75 uzoraka i 383 značajke. Dimenzije konačnog skupa podataka za testiranje su (30, 383), što znači da skup sadrži 30 uzoraka i isti broj značajki.

Ovi skupovi podataka su zatim spremljeni u .csv format za daljnju upotrebu. Format .csv je kompatibilan s većinom alata za obradu podataka i strojno učenje.

## 7. Klasifikacija afektivnih poremećaja na temelju izdvojenih značajki

### 7.1. Stablo odluke

Nakon što smo optimizirali modele klasifikacije, testirali smo ih na neviđenim podacima kako bismo procijenili njihovu generalizacijsku sposobnost.

Za model stabla odluke, odabrano je 4 značajke za klasifikaciju: higuchijska fraktalna dimenzija za beta valove u prvom kanalu EEG signala (Higuchi\_beta\_ch0), sevcikova fraktalna dimenzija za delta valove u šesnaestom kanalu EEG signala (sevcik\_delta\_ch16), higuchijska fraktalna dimenzija za gamma valove u šestom kanalu EEG signala (higuch\_gamma\_ch6), petrosianova fraktalna dimenzija za beta valove u šesnaestom kanalu EEG signala (petrosian\_beta\_ch16).

Nakon ispitivanja različitih vrijednosti hiperparametara za model stabla odluke, odabrani su najbolji hiperparametri za optimizaciju modela.

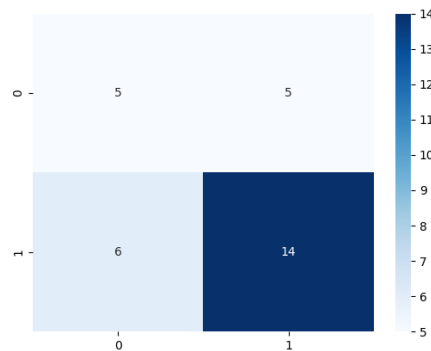
Maksimalna dubina stabla (engl. max depth) ispitivana je s mogućim vrijednostima None, 5, 10, 15, 20, a optimalna vrijednost pronađena za ovaj model je None, što znači da dubina stabla nije ograničena.

Minimalan broj uzoraka potreban za razdvajanje čvora (engl. min samples split) testiran je s mogućim vrijednostima 2, 5, 10. Kroz ovaj proces otkrili smo da je optimalna vrijednost za ovaj hiperparametar 2, što znači da je potrebno najmanje dva uzorka da bi se čvor razdvojio.

Konačno, minimalan broj uzoraka potreban za formiranje listova (engl. min samples leaf) ispitivao se s mogućim vrijednostima 1, 2, 5. Za ovaj model, optimalna vrijednost odabrana za ovaj hiperparametar je 5, što znači da svaki list u stablu mora sadržavati najmanje pet uzoraka.

Srednja točnost modela na skupu za treniranje kroz unakrsnu validaciju sa 5 preklopa iznosila je 78,67%. Na skupu za testiranje, model je postigao točnost od 63,33%.

Matrica konfuzije pokazuje da je model stabla odluke ispravno klasificirao 5 pri-



**Slika 7.1:** Matrica konfuzije za model stabla odluke

mjera klase 0 (zdravi ispitanici) od 10 i 14 primjera klase 1 (ispitanici s afektivnim poremećajem) od 20. Međutim, model je pogrešno klasificirao 5 primjera klase 0 kao klasu 1 i 6 primjera klase 1 kao klasu 0.

## 7.2. K-najbližih susjeda (KNN)

Za model KNN koristili smo 21 značajku iz skupa podataka. U procesu optimizacije modela klasifikacije KNN, ispitivali smo različite vrijednosti hiperparametara kako bismo odabrali najbolje.

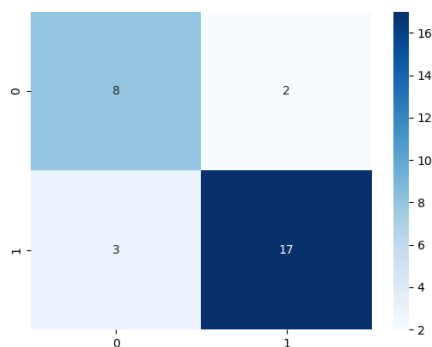
Broj susjeda (engl. *n neighbors*) koji se uzimaju u obzir prilikom donošenja odluke o klasifikaciji ispitivali smo s mogućim vrijednostima 3, 5, 7, 10. Kroz ovaj postupak, otkrili smo da optimalna vrijednost za ovaj hiperparametar za naš model iznosi 10, što znači da se za klasifikaciju uzima u obzir 10 najbližih susjeda.

Način na koji se težine dodjeljuju susjedima prilikom donošenja odluke o klasifikaciji (engl. *weights*) ispitivali smo s mogućim vrijednostima *uniform* i *distance*. Optimalna vrijednost za naš model je *uniform*, što znači da svi susjedi imaju jednaku težinu prilikom donošenja odluke o klasifikaciji.

Metriku koja se koristi za izračunavanje udaljenosti između uzoraka (engl. *metric*) ispitivali smo s mogućim vrijednostima *euclidean* i *manhattan*. Kroz ovaj proces, otkrili smo da je za naš model optimalna metrika *manhattan*, što znači da se za izračunavanje udaljenosti koristi Manhattanova udaljenost.

Srednja točnost modela na skupu za treniranje kroz unakrsnu validaciju sa 5 preklopa iznosila je 82,67%. Na skupu za testiranje, model je postigao točnost od 83,33%.

Matrica konfuzije pokazuje da je KNN model ispravno klasificirao 8 primjera klase 0 od 10 i 17 primjera klase 1 od 20. No, model je pogrešno klasificirao 2 primjera klase



**Slika 7.2:** Matrica konfuzije za Knn model

0 kao klasu 1 i 3 primjera klase 1 kao klasu 0.

Rezultati modela stabla odluke i modela KNN otkrivaju značajne razlike koje zaslužu daljnju analizu.

Broj značajki koje koristi model stabla odluke je znatno manji u usporedbi s brojem značajki koje koristi model KNN. Ova razlika može sugerirati da model stabla odluke može biti manje osjetljiv na šum u podacima, budući da koristi manje značajki za donošenje odluka. Međutim, istovremeno, postoji mogućnost da model stabla odluke može propustiti neke važne informacije koje su sadržane u značajkama koje nisu odabrane.

### 7.3. Usporedba modela

U pogledu točnosti, model KNN postigao je veću točnost na skupu za testiranje u usporedbi s modelom stabla odluke. Ovaj rezultat može sugerirati da model KNN bolje generalizira na neviđenim podacima. Međutim, važno je naglasiti da je točnost samo jedna od metrika za ocjenjivanje performansi modela.

Ukupno gledano, iako oba modela pokazuju obećavajuće rezultate, model KNN pokazao je bolje performanse u ovom kontekstu, no ne smijemo isključiti mogućnost prenaučivosti modela.

# ZAKLJUČAK

Ovaj istraživački rad predstavlja sveobuhvatno istraživanje o primjeni modela strojnog učenja za klasifikaciju afektivnih poremećaja koristeći EEG signale. Koristimo fraktalnu dimenziju kao značajku EEG signala, koja se pokazala kao značajan faktor u procesu klasifikacije. Istraživanje je provedeno koristeći skup podataka iz Klinike za psihijatriju Vrapče, koji je podijeljen na skup za treniranje i skup za testiranje. Skup za treniranje sastojao se od 50 pacijenata s afektivnim poremećajima i 25 zdravih osoba, dok je skup za testiranje uključivao 20 pacijenata i 10 zdravih osoba.

Za proces klasifikacije koristila su se dva modela strojnog učenja, konkretno stablo odluke i KNN. Modeli su trenirani i testirani koristeći izdvojene značajke iz EEG signala. Model stabla odluke postigao je prosječnu točnost od 78,67% na skupu za treniranje kroz unakrsnu validaciju s 5 preklopa i pokazao točnost od 63,33% na skupu za testiranje. S druge strane, model KNN postigao je prosječnu točnost od 82,67% na skupu za treniranje kroz unakrsnu validaciju s 5 preklopa i pokazao veću točnost od 83,33% na skupu za testiranje.

Istraživanje je pokazalo potencijal strojnog učenja u području medicinske dijagnostike, posebno u klasifikaciji afektivnih poremećaja. Korištenje EEG signala i fraktalnih dimenzija kao značajki pružilo je novi pristup problemu. Rezultati ovog istraživanja mogu otvoriti put za daljnja istraživanja i poboljšanja u području strojnog učenja i medicinske dijagnostike. Performanse modela KNN, posebno, sugeriraju da bi mogao biti obećavajući alat za klasifikaciju afektivnih poremećaja koristeći EEG signale. Budući rad bi mogao istražiti druge tehnike strojnog učenja ili metode izdvajanja značajki kako bi se dalje poboljšala performansa klasifikacije.

# LITERATURA

- [1] NHS. Electroencephalogram (eeg). nhs.uk, lipanj 2020. Pristupljeno 13. svibnja 2023.
- [2] Gonzalo M Rojas. The-10-20-international-system. researchgate.net, kolovoz 2017. Pristupljeno 25. svibnja 2023.
- [3] NeuroHealth Associates. Brainwave frequencies. nhahealth.com, siječanj 2022. Pristupljeno 23. svibnja 2023.
- [4] Wise Wunu. 5 types of brain waves frequencies: Gamma, beta, alpha, theta, delta. mentalhealthdaily.com, siječanj 2018. Pristupljeno 3. svibnja 2023.
- [5] Jaime Moreno. Brainwaves. researchgate.net, svibanj 2020. Pristupljeno 30. svibnja 2023.
- [6] Sara Brown. Machine learning explained. mitsloan.mit.edu, travanj 2021. Pristupljeno 12. svibnja 2023.
- [7] Julianna Delua. Supervised vs. unsupervised learning. ibm.com, ožujak 2015. Pristupljeno 19. svibnja 2023.
- [8] Board Infinity. Machine learning in python: Supervised vs unsupervised learning. medium.com, travanj 2020. Pristupljeno 21. svibnja 2023.
- [9] Amal Joby. Cross validation explained. learn.g2.com, listopad 2021. Pristupljeno 1. svibnja 2023.
- [10] Alexandre Rosseto Lemos. Understanding cross validation. towardsdatascience.com, travanj 2022. Pristupljeno 30. svibnja 2023.
- [11] Ajay Kulkarni. Confusion matrix. sciencedirect.com, 2020. Pristupljeno 28. svibnja 2023.



- [12] James Thorn. Decision tree. [towardsdatascience.com](https://towardsdatascience.com), ožujak 2020.
- [13] Neelam Tyagi. Gini index. [medium.com](https://medium.com), travanj 2022.
- [14] Onel Harrison. Machine learning basics with the k-nearest neighbors algorithm. [towardsdatascience.com](https://towardsdatascience.com), kolovoz 2018. Pristupljeno 15. svibnja 2023.
- [15] [towardsdatascience](https://towardsdatascience.com). Confusion matrix. [towardsdatascience.com](https://towardsdatascience.com), kolovoz 2019. Pristupljeno 26. svibnja 2023.
- [16] Sayak Paul. Parameter optimization in machine learning models. [datacamp.com](https://datacamp.com), kolovoz 2018. Pristupljeno 20. svibnja 2023.
- [17] Johnny Johnson Nancy Garcia Leah Griffith, Veronica Holbrook. Fractal dimension. [math.libretexts.org](https://math.libretexts.org), rujan 2021. Pristupljeno 9. svibnja 2023.
- [18] Slađana Z. Spasić Srđan Kesić. A new approach to estimate fractal dimension: Higuchi fractal dimension. [sciencedirect.com](https://sciencedirect.com), rujan 2016. Pristupljeno 24. svibnja 2023.
- [19] Michael J Katz. Katz fractal dimension. [sciencedirect.com](https://sciencedirect.com), 1988. Pristupljeno 6. lipnja 2023.
- [20] S. Fu, Y.-B Li, F. Ye, and Z.-G Gao. Fast blind spectrum sensing using sevcik fractal dimension in frequency domain. 05 2014. Pristupljeno 17. svibnja 2023.
- [21] Carlos Sevcik. A fast algorithm for sevcik fractal dimension calculation. [arxiv.org](https://arxiv.org), ožujak 2010. Pristupljeno 28. svibnja 2023.
- [22] Raphael vallat. Petrosian fractal dimension. [raphaelvallat.com](https://raphaelvallat.com), 2018. Pristupljeno 6. lipnja 2023.

## **Klasifikacija afektivnih poremećaja korištenjem fraktalne dimenzije kao značajke signala EEG-a**

### **Sažetak**

U ovom radu istražena je mogućnost klasifikacije afektivnih poremećaja koristeći fraktalne dimenzije kao značajke signala EEG-a. Korištena je baza podataka nastala u Klinici za psihijatriju Vrapče, koja uključuje podatke o pacijentima s dijagnozama F32 (depresija) i F33 (povratni depresivni poremećaj).

Nakon predobrade signala, koja je uključivala filtriranje signala u frekvencijskom rasponu od 0,1 Hz do 40 Hz i primjenu nezavisne analize komponenti (ICA) za identifikaciju artefakata, signali su podijeljeni na pet frekvencijskih raspona, svaki predstavljajući jedan moždani val.

Za svaki od 19 kanala, izračunate su četiri različite fraktalne dimenzije - Higuchijeva, Sevcikova, Katzova i Petrosianova. Ove fraktalne dimenzije su poslužile kao značajke za daljnju analizu.

Nakon što su dodane informacije o dijagnozi, dobiveni su konačni skupovi podataka za treniranje i testiranje. Skup za učenje uključuje 50 bolesnih ispitanika i 25 zdravih, dok skup za testiranje uključuje 20 bolesnih i 10 zdravih ispitanika.

Rezultati ovog rada pokazuju da je moguće koristiti fraktalne dimenzije kao značajke signala EEG-a za klasifikaciju afektivnih poremećaja.

**Ključne riječi:** Strojno učenje, EEG, klasifikacija, afektivni poremećaji, fraktalna dimenzija, značajke signala.

## **Classification of affective disorders using fractal dimension as a feature of EEG signals**

### **Abstract**

In this study, the possibility of classifying affective disorders using fractal dimensions as features of EEG signals was explored. The database used was created at the Vrapče Psychiatric Clinic, which includes data on patients with diagnoses F32 (depression) and F33 (recurrent depressive disorder).

After preprocessing the signals, which involved filtering the signals in the frequency range from 0,1 Hz to 40 Hz and applying Independent Component Analysis (ICA) to identify artifacts, the signals were divided into five frequency ranges, each representing a brain wave.

For each of the 19 channels, four different fractal dimensions were calculated - Higuchi's, Sevcik's, Katz's, and Petrosian's. These fractal dimensions served as features for further analysis.

After adding diagnostic information, the final datasets for training and testing were obtained. The training set includes 50 ill examinees and 25 healthy ones, while the testing set includes 20 ill and 10 healthy examinees.

The results of this study show that it is possible to use fractal dimensions as features of EEG signals for the classification of affective disorders.

**Keywords:** Machine learning, EEG, classification, affective disorders, fractal dimension, signal features.