

# Predikcija poremećaja iz spektra autizma korištenjem značajki strukturnih slika magnetske rezonancije mozga

---

Barić, Lana

Undergraduate thesis / Završni rad

2024

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:168:249974>

*Rights / Prava:* [In copyright](#) / [Zaštićeno autorskim pravom](#).

*Download date / Datum preuzimanja:* **2025-03-14**



*Repository / Repozitorij:*

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repozitory](#)



SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1270

**PREDIKCIJA POREMEĆAJA IZ SPEKTRA AUTIZMA  
KORIŠTENJEM ZNAČAJKI STRUKTURNIH SLIKA  
MAGNETSKE REZONANCIJE MOZGA**

Lana Barić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1270

**PREDIKCIJA POREMEĆAJA IZ SPEKTRA AUTIZMA  
KORIŠTENJEM ZNAČAJKI STRUKTURNIH SLIKA  
MAGNETSKE REZONANCIJE MOZGA**

Lana Barić

Zagreb, lipanj 2024.

## ZAVRŠNI ZADATAK br. 1270

Pristupnica: **Lana Barić (0036538219)**  
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo  
Modul: Računarstvo  
Mentorica: doc. dr. sc. Jelena Božek

Zadatak: **Predikcija poremećaja iz spektra autizma korištenjem značajki strukturnih slika magnetske rezonancije mozga**

### Opis zadatka:

Analizom slika magnetske rezonancije (MR) mozga mogu se odrediti značajke za predikciju poremećaja iz spektra autizma, a dobivene značajke i rezultati mogu doprinijeti odlukama o dijagnozi, prognozi i liječenju autizma. U radu je potrebno opisati način snimanja mozga magnetskom rezonancijom. Navesti osnovne korake analize strukturnih MR slika mozga. Cilj rada je korištenjem algoritma strojnog učenja provesti predikciju poremećaja iz spektra autizma na temelju značajki dobivenih iz strukturnih MR slika mozga. Provesti validaciju i testiranje algoritma na javno dostupnim podacima. Prikazati i ocijeniti ostvarene rezultate.

Rok za predaju rada: 14. lipnja 2024.



## Sadržaj

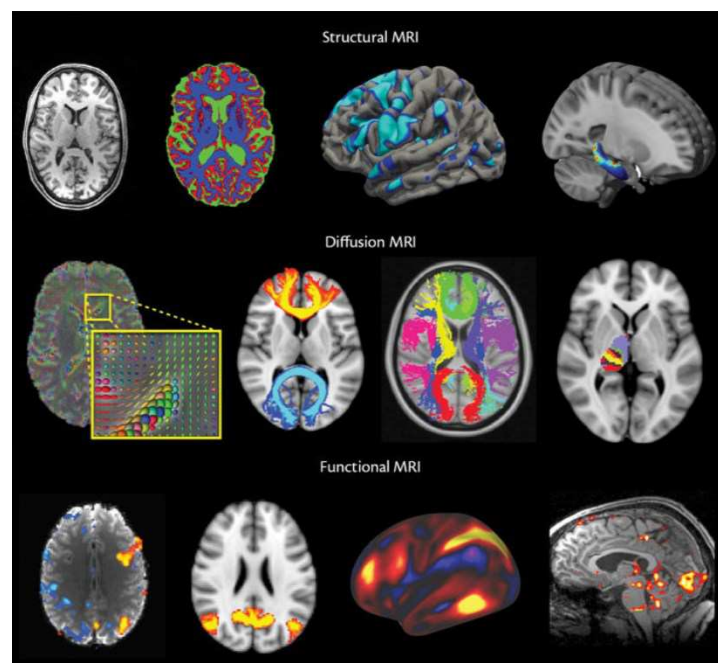
Uvod .....	1
1. Magnetska rezonancija .....	2
1.1. Fizika iza magnetske rezonancije .....	3
1.1.1. Magnetsko polje (B0 polje) .....	4
1.1.2. Rezonancija .....	4
1.1.3. Snimanje .....	5
2. Strukturni MRI .....	6
2.1. Strukturni cjevovod .....	8
3. Materijali i metode.....	11
3.1. Skup podataka .....	11
3.2. Predikcija strojnim učenjem .....	14
3.2.1. Logistička regresija.....	14
3.3. Provedba predikcije autizma .....	15
4. Rezultati.....	17
4.1. Usporedba rezultata .....	23
Zaključak .....	24
Literatura .....	25
Sažetak.....	26
Summary.....	27
Skraćenice.....	28

# Uvod

Autizam je kompleksan neurobiološki poremećaj ljudskog mozga okarakteriziran različitim oblicima oštećenja socijalne interakcije, komunikacije i ponašanja. Otkriva se u djetinjstvu i (obično) prati osobu tijekom cijelog života [1]. Uzrok je nepoznat, ali smatra se da razni genetski, neurološki i okolišni faktori zajedno igraju ulogu u razvoju autizma. Rana dijagnoza i pravovremena intervencija izrazito su važne kako bi osobe s autizmom dobile najbolju moguću pomoć. Kroz godine istraživanja, postalo je jasno da razumijevanje ovog poremećaja zahtijeva interdisciplinarni pristup, uključujući i napredne tehnologije poput strojnog učenja. Algoritmi strojnog učenja idealni su za obradu velike količine informacija u svrhu identificiranja uzoraka i poveznica koje ljudski um možda ne bi mogao detektirati. Stoga oni omogućuju preciznije dijagnoze te praćenje napretka osoba s autizmom. Magnetska rezonancija (MR) mozga je napredna i neinvazivna medicinska tehnika dijagnostike koja pruža detaljne i jasne slike mozga i njegovih struktura. Postoji nekoliko vrsta MR slika mozga, a u ovom radu fokus je na strukturne slike magnetske rezonancije mozga. Strukturno oslikavanje magnetske rezonancije (sMRI) pruža detaljan uvid u anatomiju mozga. Analizom slika mozga dobivenih iz sMRI-a, uz pomoć algoritama strojnog učenja, mogu se identificirati karakteristične promjene u strukturi mozga koje su povezane s autizmom i napraviti predikcija tog poremećaja.

# 1. Magnetska rezonancija

Magnetska rezonancija mozga napredna je tehnika medicinske dijagnostike koja koristi snažne magnete, radiovalove i računalnu tehnologiju u svrhu stvaranja detaljnih slika mozga i njegovih struktura. Liječnici ju koriste kako bi pregledali unutrašnjost mozga bez invazivnih procedura. Postoji široki raspon vrsta slika koje se mogu dobiti na MR skeneru, a različite vrste bilježe različite informacije o mozgu te se nazivaju modalitetima. Tri modaliteta koja se najčešće koriste za oslikavanje mozga su strukturno, funkcionalno i difuzijsko snimanje (Slika 1.1). Strukturno snimanje daje informacije o osnovnim anatomskim strukturama u mozgu (npr. granice moždane kore te granice drugih struktura poput hipokampusa). Funkcionalno snimanje pruža informacije o aktivnosti neurona u mozgu, kao odgovor na specifične podražaje ili na zadatke koje istraživač daje pacijentu ili u vezi sa spontanom aktivnošću neurona sudionika. Difuzijsko snimanje daje informacije o mikrostrukтури i anatomskoj povezanosti unutar mozga i leđne moždine. Često se u istraživanjima istovremeno koristi nekoliko modaliteta jer se jedan modalitet može koristiti za nadopunu drugoga.

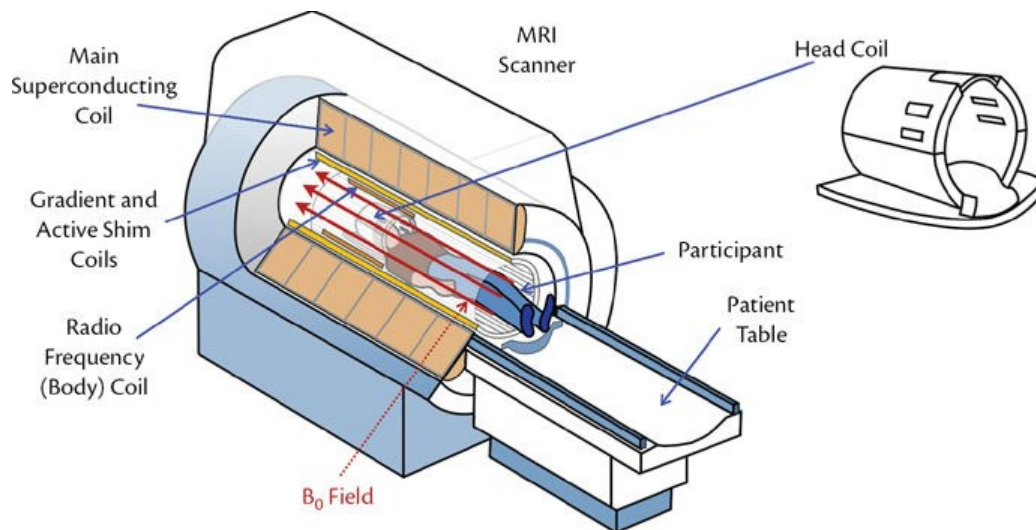


Slika 1.1 Tri glavna modaliteta MR slika mozga [2]



## 1.1. Fizika iza magnetske rezonancije

MR se temelji na principu da određene atomske jezgre, koje se još nazivaju „spinovi“, djeluju poput malih štapićastih magneta i reagiraju s magnetskim poljima na način koji omogućuje mjerenje i manipulaciju njihovim magnetskim stanjem. Budući da se mjeri i manipulira samo magnetsko stanje, ne dolazi do nikakve štete jezgara i nema posljedica na bilo koji biološki proces u kojem sudjeluju molekule u kojima su te jezgre. Prilikom snimanja mozga, prvenstveno se cilja na jezgre vodika unutar molekula vode jer su obilne, ali i jezgre vodika u masnim molekulama doprinose signalima te se pojavljuju u nekim vrstama skeniranja. Kako bi se stvorilo i izmjerilo magnetsko polje, MR skener koristi skup različitih zavojnica napravljenih od električno provodljive žice (Slika 1.2). Koriste se zavojnice jer kada električna struja prolazi kroz nju, stvara se magnetsko polje, a promjena magnetskih polja unutar zavojnice inducira električnu struju u njoj. S obzirom na to da se jezgre vodika u molekulama vode ponašaju poput malih štapićastih magneta, mogu se manipulirati magnetskim poljima koja se stvaraju kada struja prolazi tim zavojnicama. Osim toga, jezgre samostalno proizvode mala magnetska polja koja su dovoljna za stvaranje malih električnih struja u nekima od tih zavojnica, a posebno u zavojnici za glavu koja je najbliža.



Slika 1.2 Shema MR skenera [2]

### 1.1.1. Magnetsko polje (B0 polje)

Da bi došlo do interakcije s vodikovim jezgrama, potrebno je vrlo jako statičko magnetsko polje jer samo u takvom polju maleni magneti iz jezgara imaju tendenciju usmjeravanja u istom smjeru (duž glavnog polja). Ako ne pokazuju u istom smjeru, zbroj njihovih magnetskih polja će biti (gotovo) nula i ne može se detektirati. To jako magnetsko polje, poznatije kao B0 polje, uspostavljeno je velikom supravodljivom zavojnicom koja je uvijek uključena i hladi se tekućim helijem. Baš to polje definira snagu skenera, koja se izražava u jedinicama Tesla (T). Na primjer, 3T skener je onaj gdje je B0 polje 3T. MR skener sadrži „*shim*“ zavojnice koje stvaraju polja koja pokušavaju poništiti neujednačenosti u glavnom B0 polju. Bitno je minimizirati ikakve nejednolikosti u B0 polju kako bi se izbjegli artefakti na slici pa se zato na početku svakog skeniranja primjenjuje postupak „*shimming*“.

### 1.1.2. Rezonancija

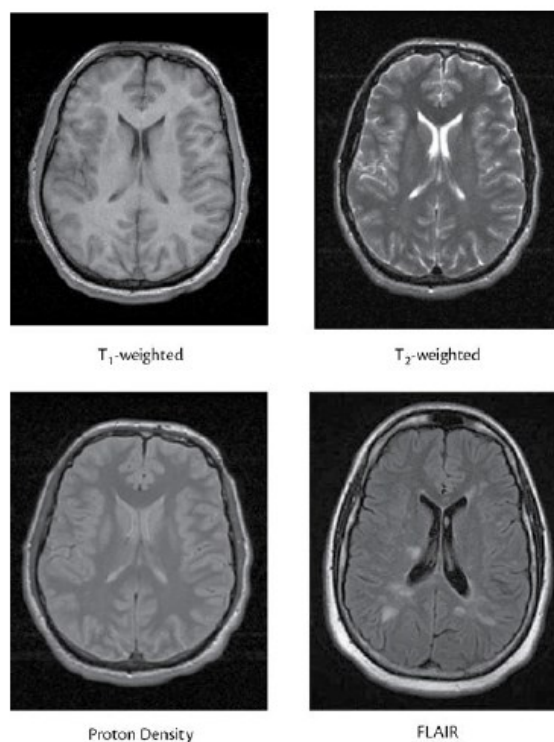
Ključna značajka jezgri vodika u snažnom polju je da mali štapić magneti rotiraju oko osi B0 polja, ali na takav način da je frekvencija rotacije proporcionalna snazi vanjskog magnetskog polja. Ta rotacija stvara oscilirajuća polja (B1 polja) koja se mogu detektirati i manipulirati izvana pri poznatoj frekvenciji. Ta frekvencija – rezonantna frekvencija – je ona na koju su jezgre podešene, a to omogućava da se njima manipulira odašiljanjem na toj frekvenciji te primanjem signala koji dolaze od njih na toj frekvenciji. Rezonantno svojstvo je ključna značajka MR-a koja nam omogućava dobivanje signala. Ta interakcija se odvija preko radio frekvencijskih (RF) zavojnica. Dva najčešća primjera su zavojnica za tijelo i zavojnica za glavu. Zavojnica za glavu se postavlja oko glave sudionika neposredno prije nego što on uđe u skener te je to jedina zavojnica koja nije ugrađena u uređaj. Glavni razlog tomu je što je korisno postaviti ju što bliže objektu koji se skenira i što postoji nekoliko tipova zavojnica koje su korisne u različitim situacijama pa su dizajnirane da budu lako zamjenjive. Svrha zavojnice je odašiljanje ili primanje signala oko rezonantne frekvencije jezgri – B1 polja. Zavojnica za tijelo je ugrađena u skener i obavlja istu funkciju kao zavojnica za glavu. Ona je obično mnogo dalje od sudionikove glave pa je zato manje osjetljiva za primanje slabijih signala koje emitiraju jezgre vodika unutar mozga. Zbog toga se ne koristi za primanje signala u oslikavanju mozga, ali može pružiti snažno i ujednačeno polje za manipulaciju stanja magnetizacije jezgri vodika pa je preferirana zavojnica za odašiljanje.

### 1.1.3. Snimanje

Rezonantna svojstva vrijede za svaku pojedinu jezgru vodika što znači da se opaženi MR signal sastoji od doprinosa svih jezgri unutar cijevi skenera zbrojenih zajedno. Da bi se odredilo odakle signali dolaze (s koje lokacije unutar mozga) potrebno je izdvojiti pojedinačne doprinose tom zbrojenom signalu. To se radi na temelju činjenice da rezonantna frekvencija, a time i frekvencija signala, ovisi o jačini polja. S obzirom na to da je moguće izdvojiti doprinose različitih frekvencija u kombiniranom signalu, lokacija signala se može odrediti ako se frekvencija može povezati s lokacijom. Kako bi se ta povezanost ostvarila, dodaju se dodatna, pažljivo kontrolirana, magnetska polja koja variraju s lokacijom. Ta polja se dodaju prilikom provedbe mjerenja signala, tako da signali s različitih lokacija imaju različite i poznate frekvencije. To omogućuje mjerenje jačine signala na svakoj frekvenciji te se na temelju toga izračunava koliko signala je došlo s određene prostorne lokacije. Dodatna polja se nazivaju gradijentna polja i stvaraju ih tri različite gradijentne zavojnice u skeneru. Primjenom gradijenta polje varira s lokacijom, na primjer jedna gradijentna zavojnica može stvoriti polje koje je slabo na dnu mozga i jako na vrhu, linearno varirajući između. Postojanje polja koje se mijenja s lokacijom dovodi do toga da signali iz jezgri vodika imaju frekvenciju koja ovisi o njihovoj lokaciji, a takvo mjerenje frekvencijskog sadržaja primljenih signala omogućuje određivanje njihove lokacije i formiranje slike. Bilo koje nehomogenosti u  $B_0$  polju ili nesavršenosti u gradijentnim poljima dovode do iskrivljenja u konačnoj slici. Ta iskrivljenja su obično pogrešne lokacije signala jer kada polje nije na pravoj vrijednosti, nije ni frekvencija signala pa je lokacija pogrešno izračunata. Osim toga, velike nehomogenosti u  $B_0$  polju mogu uzrokovati gubitak signala, dok nehomogenosti u bilo kojem odašiljanom ili primljenom  $B_1$  polju rezultiraju promjenama intenziteta slike. To se manifestira kao svjetlija ili tamnija područja na slikama, a te posljedice nehomogenosti se rješavaju metodama analize.

## 2. Strukturni MRI

Strukturne slike magnetske rezonancije (sMRI) u oslikavanju mozga obično se koriste za prikaz grube anatomije mozga: uglavnom sive tvari, bijele tvari i cerebrospinalne tekućine. Budući da se sMRI koristi za identifikaciju detalja pojedinačne anatomije i neophodan je za precizno poravnanje mozgovih različitih ispitanika, bitan je dio grupnog istraživanja te se podaci dobiveni iz njega koriste uz one dobivene drugim vrstama magnetske rezonancije. Vrijeme skeniranja strukturnih slika magnetske rezonancije mozga može varirati od nekoliko minuta pa sve do 15 minuta, ali uobičajeno je između 3 i 5 minuta. T<sub>1</sub> mjerene slike su najčešće vrste strukturnih slika, ali postoje različite varijante poput T<sub>2</sub> mjerene, PD (engl. *Proton density*) te FLAIR (Slika 2.1). Razlog postojanja više tipova strukturnih slika je što svaki tip ističe drugačije aspekte prisutnih tkiva, pružajući vrlo vrijedan način istraživanja anatomije in vivo.



Slika 2.1 Primjeri vrsta strukturnih MR slika [2]

Magnetna svojstva jezgre vodikovih atoma u molekulama vode određuju MR signal, a za strukturno snimanje važne su samo tri ključne karakteristike: gustoća protona i dva relaksacijska procesa, opisana T<sub>1</sub> i T<sub>2</sub> relaksacijskim konstantama. Ove karakteristike određene su lokalnim mikroskopskim okruženjem molekula vode i različite su unutar različitih vrsta tkiva, što omogućuje dobivanje MR slika koje imaju izražen kontrast između

tih tkiva. Gustoća protona je jednostavno svojstvo – ono predstavlja gustoću protona koji doprinose MR signalu. Relaksacijske konstante, T1 i T2, određene su mnogim drugim aspektima molekula vode i njihovim okruženjem, poput geometrije i rasporeda susjednih stanica, okolnih molekula i još mnogo toga. Kao posljedica toga, moguće je dizajnirati MR sekvence koje su osjetljive na te promjene, što omogućava velike razlike u intenzitetu između različitih tkiva, posebno u usporedbi s drugim modalitetima koji nisu MR.

Međutim, poznavanje vrijednosti ovih triju fizikalnih parametara (gustoće protona i dvaju relaksacijskih vremena), ili numeričke vrijednosti intenziteta signala koji su njima određeni, nije dovoljno informacija da bi se mogle utvrditi sve biološke karakteristike tkiva. Stoga signal koji se mjeri na strukturnoj slici, tj. intenzitet slike, ističe strukturne granice od interesa (između tkiva), ali djeluje samo kao zamjena za tip tkiva, umjesto da pruža kvantitativne informacije o specifičnim biološkim svojstvima.

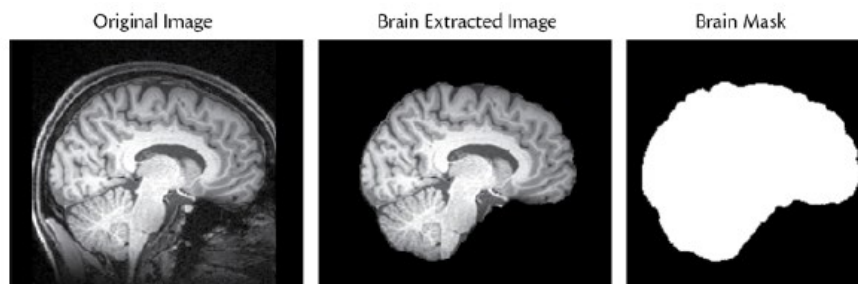
Za strukturne slike najvažnija je vidljivost granica između tkiva kako bi prikazale grubu anatomiju; a značajne razlike u intenzitetu između tkiva ključne su za to. Kontrast na slici definira se kao razlika u intenzitetu kroz granice između dvaju tkiva ili struktura od interesa. Vrlo bitna je veličina kontrasta u odnosu na veličinu šuma, a to se mjeri omjerom kontrasta i šuma (CNR, engl. *Contrast-to-noise ratio*). Za strukturno snimanje, CNR je jedna od najvažnijih karakteristika slike, zajedno s prostornom rezolucijom.

Moguće je dobiti različite tipove strukturnih slika manipulacijom utjecaja svakog od tri parametra. No gustoća protona je najosnovnije svojstvo od ta tri, jer gustoća protona od nule znači da nema molekula vode koje bi doprinijele signalu, pa je signal uvijek proporcionalan gustoći protona, bez obzira na vrstu slike. Promjenom vremena različitih magnetskih polja primijenjenih od strane skenera, intenzitet slike postaje osjetljiviji na relaksacijska svojstva. Upravo ta vremena, zajedno sa svojstvima tkiva, određuju intenzitete slike, a time i kontrast. Glavni vremenski parametri koji se prilagođavaju u strukturnim sekvencama impulsa su vrijeme ponavljanja (TR, engl. *Repetition time*) i vrijeme odjeka (TE, engl. *Echo time*).

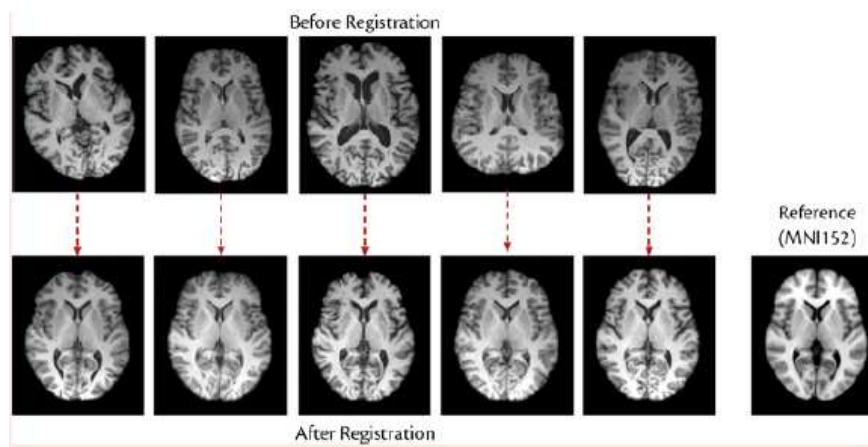
Najčešća prepreka svim MRI (engl. *Magnetic Resonance Imaging*) modelima je šum, a u strukturnim slikama CNR je najbitniji. Također, jedno specifično ograničenje sMRI-a je što ne prikazuje kost na slikama zbog toga što tamo nema dovoljno vode da se napravi upotrebljiv signal. Stoga lubanja izgleda jednako tamno kao i zrak pa se ne može raspoznati na primjer zrak u sinusima od okolnih kostiju ili frakture u lubanji.

## 2.1. Strukturni cjevovod

Cjevovod (engl. *Pipeline*) je definirani niz koraka analize koji se izvode jedan za drugim, u utvrđenom redosljedu. Postoje dva cilja strukturnih *pipeline*-a: nadodati nužne informacije koje bi podržale druge analize te analizirati anatomiju. Neki koraci su jednaki za oba cilja, ali u nekima se ipak razlikuju, a bit će navedeni u nastavku. Najčešće je za obradu i analizu potrebno maknuti strukture koje nisu dio mozga ili kreirati masku mozga (Slika 2.2). Taj korak pomaže registraciji. Registracija je proces poravnavanja slika koji se koristi da se poravnaju dvije slike od različitih sudionika ili slika sudionika s predloškom (Slika 2.3).

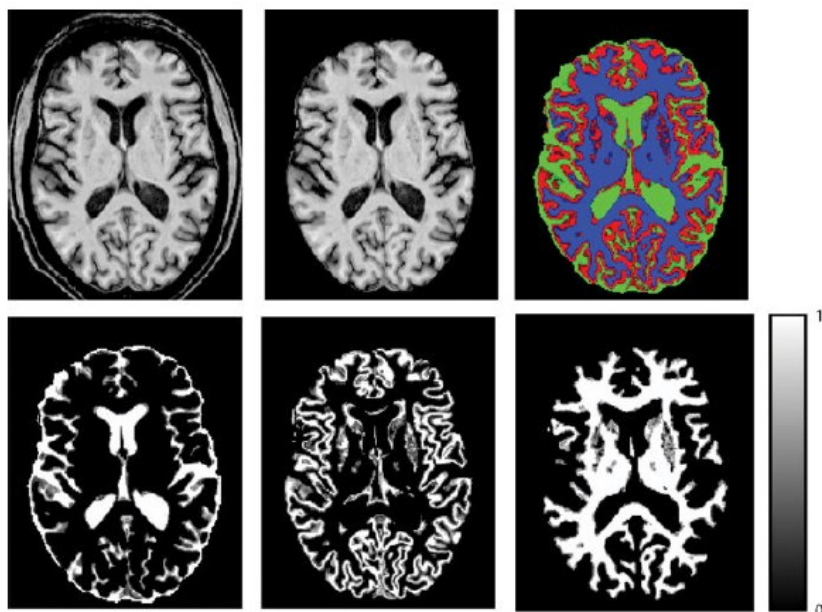


Slika 2.2 Ilustracija ekstrakcije mozga [2]



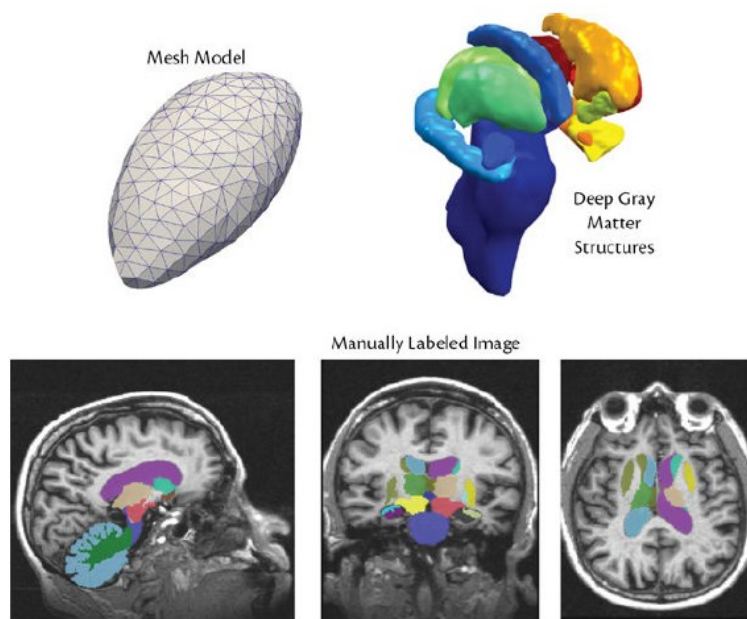
Slika 2.3 Primjer registracije [2]

Sljedeći korak u analizi je segmentacija. Postoje dvije glavne vrste segmentacije: segmentacija prema tipu tkiva i strukturna segmentacija. Obje dodjeljuju oznake različitim lokacijama na pojedinačnoj slici prema anatomskom sadržaju na toj lokaciji. Kod segmentacije prema tipu tkiva (Slika 2.4), oznake su obično siva tvar, bijela tvar, cerebrospinalna tekućina (CSF) ili mješavina ovih tkiva. Intenzitet je zaslužan za određivanje što voksel sadrži. Problem je što šum može uzrokovati da neki vokseli koji sadrže bijelu tvar imaju niži intenzitet zbog čega može doći do zabune i pomisliti da se radi o sivoj tvari i obratno.



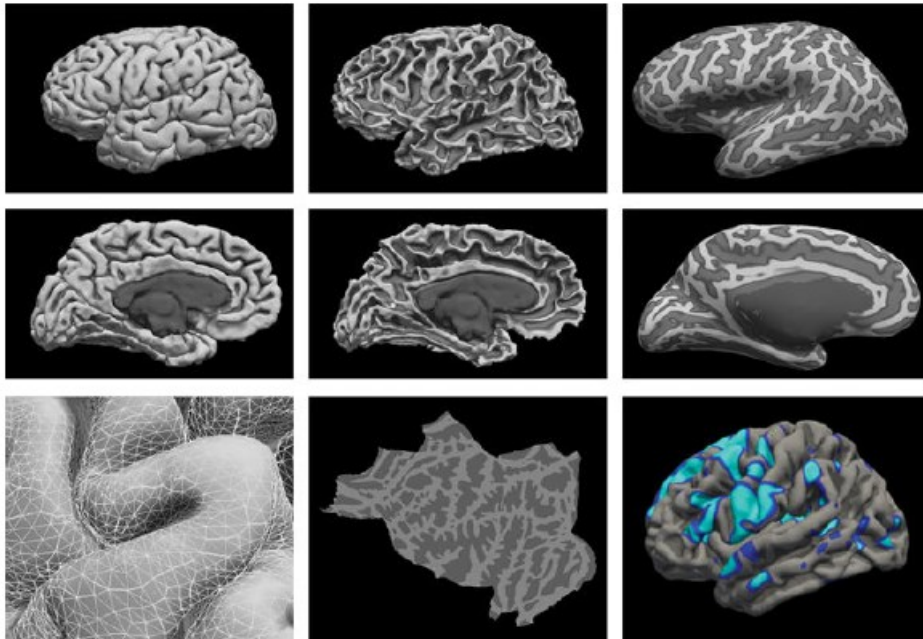
Slika 2.4 Ilustracija segmentacije prema vrsti tkiva [2]

Kod strukturne segmentacije, oznake odgovaraju specifičnijim anatomskim strukturama, poput hipokampusa ili moždane kore. Takva segmentacija zahtijeva više informacija od samih intenziteta vokseli i odnosa u susjedstvu. Potrebno je imati neki oblik anatomskog znanja o tome gdje se te strukture nalaze i kakvog su oblika. Te informacije često se specificiraju kroz skup ručno označenih slika (Slika 2.5) i mogu se koristiti za izradu eksplicitnog modela oblika i izgleda strukture u mozgu.



Slika 2.5 Ilustracija pomoći pri strukturnoj segmentaciji [2]

Segmentacija visoko naborane kortikalne površine još je izazovniji zadatak i zahtijeva visoko specijalizirane alate; dobro poznat i često korišten alat za to je paket FreeSurfer. Ta segmentacija se provodi prilagođavanjem unutarnje i vanjske površine sive tvari korteksa (mreže granica). Osim što su prikazane kortikalne površine, na slici ispod teksta (Slika 2.6) u donjem desnom dijelu se može vidjeti rezultat statističke analize provedene tako da se podaci sa svake granične točke, a ne sa svakog voksel, obrađuju zasebno.



Slika 2.6 Ilustracija kortikalne površine lijeve hemisfere [2]



### 3. Materijali i metode

Rad se temelji na izazovu opisanom u članku Traut i sur. [3]. iz kojeg je preuzeti kostur programskog koda koji je onda nadograđen.

#### 3.1. Skup podataka

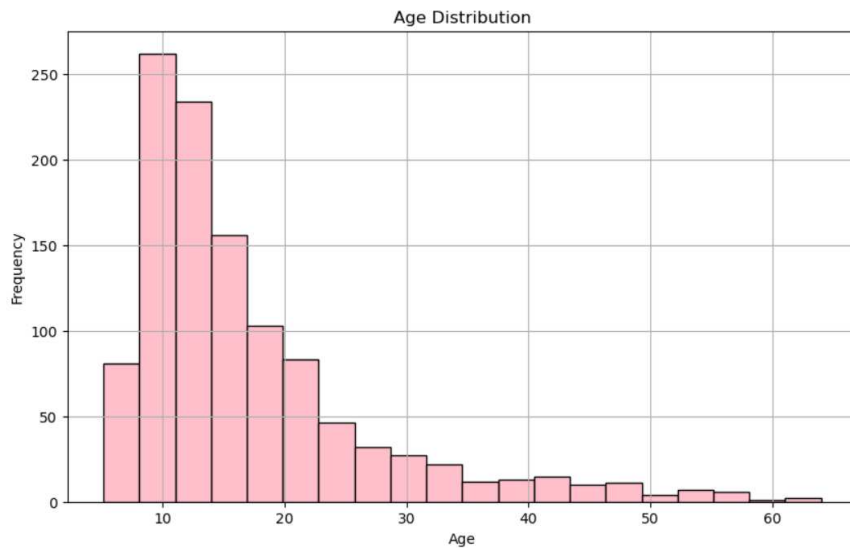
Skup podataka koji je korišten nastao je kombinacijom podataka iz javnih skupova podataka Autism Brain Imaging Data Exchange<sup>1</sup> (ABIDE) I i II . ABIDE daje otvoren pristup funkcionalnim i anatomskim MRI podacima za 2156 ispitanika. Svi MRI podaci prikupljeni su na 3 Tesla skenerima. Na svim lokacijama ispitanici su dijagnosticirani koristeći standardne ADI/ADOS alate za podršku kliničkoj procjeni. Većina ispitanika imala je ukupni IQ>75. Iz strukturnog MRI-a bile su izdvojene anatomske značajke: regionalan volumen mozga, debljina korteksa i površina korteksa. Regionalan volumen mozga odnosi se na mjerenje veličine ili volumena specifičnih područja ili regija unutar mozga. Neke od ključnih regija mozga su cerebralni korteks, subkortikalne strukture, mali mozak, moždano deblo te ventrikularni sustav. Mjerenje regionalnih volumena može dati uvid u razvoj mozga, starenje i utjecaj neuroloških poremećaja. U studijama oslikavanja mozga, tehnika MRI-a se koristi za kvantifikaciju tih volumena te se oni onda uspoređuju među populacijama kako bi se razumjele normalne i abnormalne promjene u mozgu. Debljina korteksa je udaljenost između vanjske i unutarnje površine cerebralnog korteksa. Ta debljina varira u različitim regijama mozga i pruža uvide u strukturu i funkciju mozga. U oslikavanju mozga se koristi kao i regionalni volumen, za razumijevanje razvoja i starenja mozga te učinaka neuroloških ili psihijatrijskih poremećaja. Površina korteksa se odnosi na ukupnu površinu vanjske strane cerebralnog korteksa. Na nju utječe naboranost mozga u giruse (izbočine) i sulkuse (udubine), što značajno povećava površinu u usporedbi s glatkom površinom mozga. Površina mozga je povezana s kognitivnim sposobnostima i funkcijom mozga. Kao i kod debljine korteksa, površina korteksa varira ovisno o regiji. Za dobivanje tih značajka, podaci su bili obrađeni pomoću standardnih alata za oslikavanje mozga: Freesurfer, FSL i AFNI.

Za potrebe ovog rada, korišten je javni set podataka dan od strane autora izazova. On se sastojao od 1127 ispitanika: 536 s ASD-om i 591 kontrolnih. Skup podataka trebao bi dobro

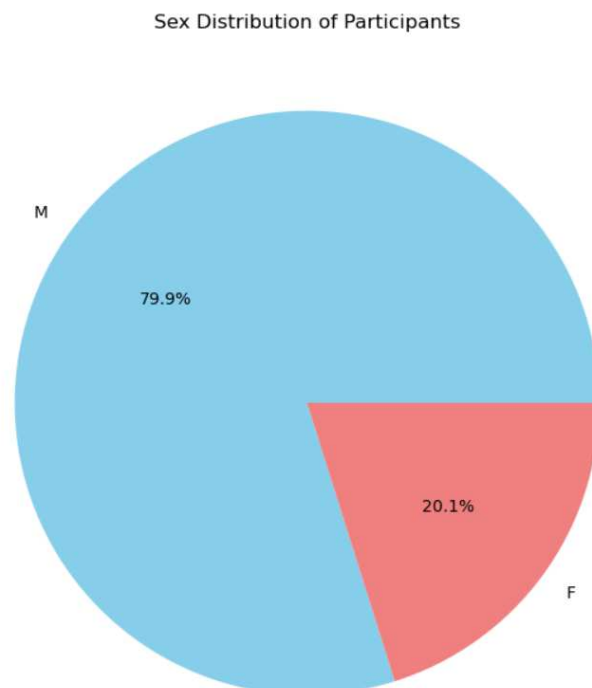
---

<sup>1</sup> [https://fcon\\_1000.projects.nitrc.org/indi/abide/](https://fcon_1000.projects.nitrc.org/indi/abide/)

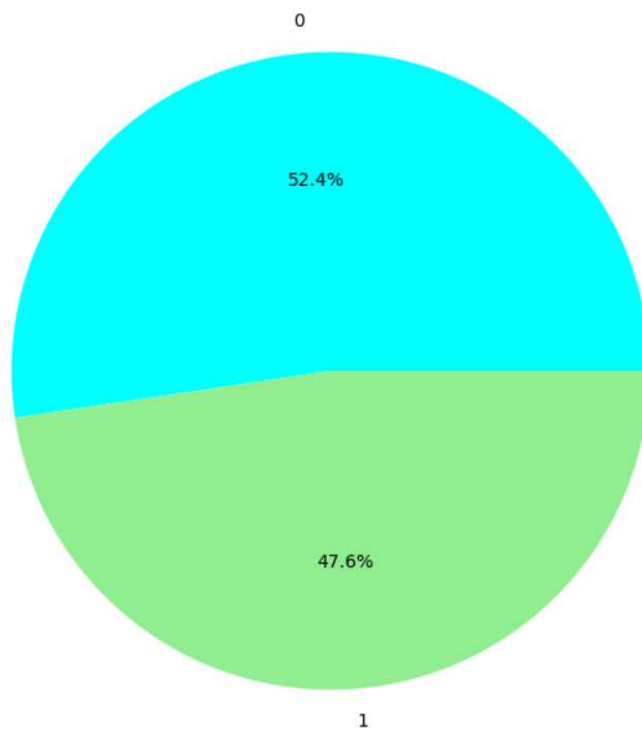
prikazati kliničku i metodološku heterogenost oslikavanja mozga ispitanika s ASD-om. Korišteni set podataka obuhvaćao je dobní raspon od 5 do 64 godine, s medijanom od 13,8 godina (Slika 3.1). Što se tiče spolne raspodjele, skup podataka se sastojao od 79,9% muškaraca te 20,1% žena (Slika 3.2). U setu je 47,6% ispitanika imalo ASD, a 52,4% ih je bilo kontrolnih (Slika 3.3). Početni set podataka nije isključio ispitanike na temelju kontrole kvalitete (Slika 3.4), ali je to onda kasnije napravljeno – za predikciju su korišteni samo oni podaci koji su bili izvrsne ili prihvatljive kvalitete (vrijednosti 1 i 2).



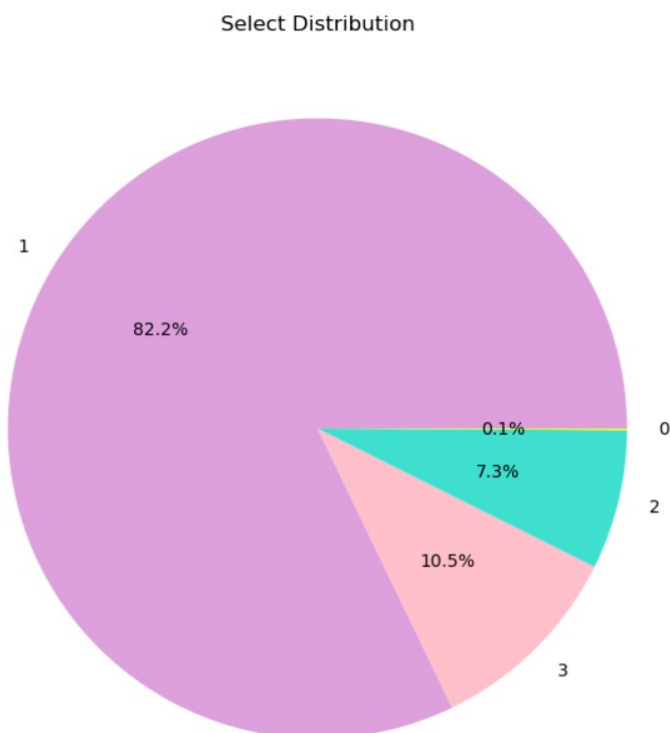
Slika 3.1 Distribucija dobnog raspona sudionika u skupu podataka



Slika 3.2 Spolna distribucija sudionika u skupu podataka



Slika 3.3 Distribucija sudionika s ASD-om na skupu podataka



Slika 3.4 Distribucija kvalitete podataka

## 3.2. Predikcija strojnim učenjem

Predikcija strojnim učenjem odnosi se na izlaz algoritma koji je uvježban na nekom skupu podataka. Algoritam generira vjerojatne vrijednosti za nepoznate varijable u svakom zapisu novih podataka. Svrha predikcije u strojnom učenju je projicirati vjerojatni skup podataka koji se odnosi na izvorne podatke [4]. U suštini, predviđanje se koristi za prilagođavanje oblika što je moguće bliže podacima. Predviđanje se može koristiti za prognoziranje budućnosti te predviđanje vjerojatnosti ishoda. Jedan od alata za predikciju je regresijska analiza koja se koristi za određivanje odnosa između dvije varijable (pojedinačna regresija) ili više od dvije varijable (višestruka regresija). Predikcija se koristi za različite namjene, od nadopunjavanja teksta na mobitelima, predviđanja ponašanja korisnika, određivanja potreba u ugostiteljstvu i industriji zabave pa sve do predviđanja zdravstvenih stanja u medicini.

### 3.2.1. Logistička regresija

Unatoč njegovom nazivu, algoritam logističke regresije je klasifikacijski algoritam koji se često koristi u praksi. Ideja logističke regresije je da se upotrijebi aktivacijska funkcija koja je slična funkciji praga, ali da je derivabilna na cijeloj svojoj domeni [5]. Za to se koristi logistička funkcija, poznatija kao sigmoidna funkcija (1).

$$\sigma(\alpha) = \frac{1}{1 + \exp(-\alpha)} \quad (1)$$

Nagib sigmoide može se kontrolirati množenjem ulaza određenim faktorom. Sigmoidna funkcija ima tri važne karakteristike: ograničava vrijednosti na (otvoreni) interval (0,1), derivabilna je, što je važno za optimizaciju te je oblikom slična funkciji praga, a to znači da će davati vrijednost blizu 1 primjerima iz jedne klase, a vrijednost blizu 0 primjerima iz druge klase.

Model logističke regresije se definira na sljedeći način (2):

$$h(\mathbf{x}; \mathbf{w}) = \sigma(\mathbf{w}^T \phi(\mathbf{x})) = \frac{1}{1 + \exp(-\mathbf{w}^T \phi(\mathbf{x}))} \quad (2)$$

Ukratko, model logističke regresije svim primjerima dodjeljuje vrijednosti iz intervala (0,1) koje se dobivaju primjenom logističke funkcije na skalarni umnožak težina i značajki. Primjeri koji su daleko od granice, imat će vrijednosti blizu 0 odnosno 1, ovisno na kojoj strani hiperravnine su, ali nikad izvan tog intervala. Vrijednost hipoteze  $h(\mathbf{x})$  se tumači kao vjerojatnost da primjer pripada pozitivnoj klasi.

Problem algoritma logističke regresije je što distribucije klasa koje su iskrivljene značajno utječu na performanse logističke regresije, pogotovo u slučajevima gdje je reprezentacija klasa neuravnotežena. Logistička regresija ima tendenciju favorizirati većinsku klasu kada jedna klasa značajno nadmašuje drugu u broju instanci. Posljedica toga je gubitak sposobnosti identificiranja i predviđanja pojava u manjinskoj klasi, što rezultira smanjenom točnošću, preciznošću i osjetljivošću [6].

### 3.3. Provedba predikcije autizma

Kako bi se došlo do zaključka je li uporaba sMRI-a u svrhu predikcije autizma prikladna, provedeno je nekoliko predikcija s različitim skupovima značajki. Najprije je početni skup podataka bio filtriran tako da uzima u obzir podatke za sudionike kod kojih je kvaliteta podataka bila izvrsna ili prihvatljiva (`filtered_df = df[df['anatomy_select'].isin([1, 2])]`). Zatim su napravljeni manji skupovi koji su sadržavali različite značajke. Promatrano je kako se mijenja točnost i preciznost predikcije autizma kada se koriste sve dane anatomske značajke mozga (207 značajki), kada razdvojimo one koje su povezane sa volumenom mozga (67 značajki) te one koje su povezane s kortikalnom debljinom i površinom (140 značajki), a zatim i imaju li utjecaja demografski podaci sudionika – spol, godine te lokacija na kojoj su snimani sudionici. Pri svakoj provedbi predikcije, za ciljnu varijablu je postavljen stupac „participants\_asd“, a ostali stupci su bili značajke. Provedeno je skaliranje podataka (Kod 3.1), a potom su oni raspodijeljeni na skup za treniranje i skup za testiranje (Kod 3.2).

```
from sklearn.preprocessing import StandardScaler
# Initialize the StandardScaler
scaler = StandardScaler()
# Fit and transform the features
scaled_features = scaler.fit_transform(features)

# Convert the scaled features back to a DataFrame for easier
handling
scaled_features_df = pd.DataFrame(scaled_features,
columns=features.columns, index=features.index)
scaled_features_df
```

Kod 3.1 Program za skaliranje podataka

```
# split X and y into training and testing sets
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test2 =
train_test_split(features, y, test_size=0.25,
random_state=16)
```

### Kod 3.2 Program za raspodjelu seta podataka

Za predikciju je korišten algoritam logističke regresije (Kod 3.3), a za evaluaciju modela korištena je matrica zabune (Kod 3.4).

```
#model development and prediction
from sklearn.linear_model import LogisticRegression

#instantiate the model
logreg = LogisticRegression(random_state=16, max_iter=500)

#fit the model with data
logreg.fit(scaled_features_df, y_train)
y_pred = logreg.predict(scaled_features_df_test)
```

### Kod 3.3 Program za predikciju pomoću logističke regresije

```
#model evaluation using confusion matrix
from sklearn import metrics
cnf_matrix = metrics.confusion_matrix(y_test2, y_pred)
cnf_matrix
```

### Kod 3.4 Program za izradu matrice zabune

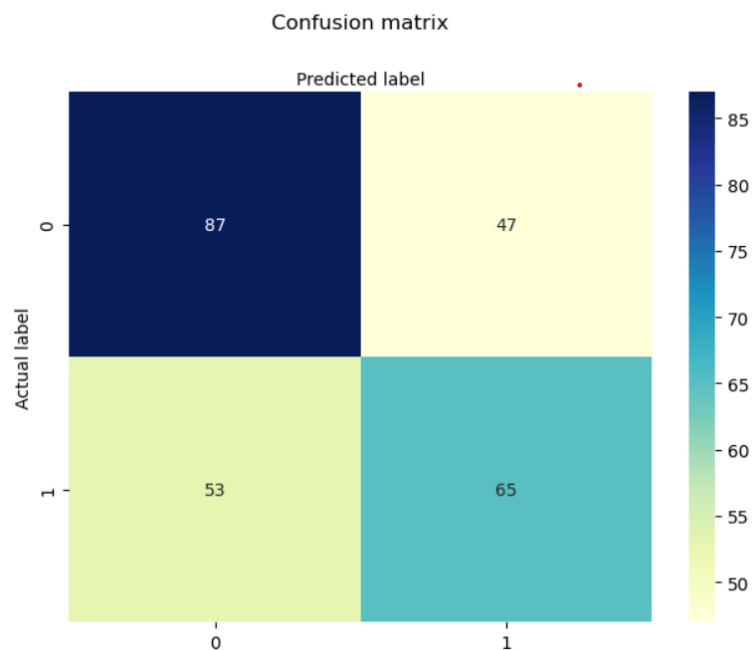
Skup podataka i Jupyter bilježnica koji su korišteni za ovaj rad, dostupni su na GitHubu.<sup>2</sup>

---

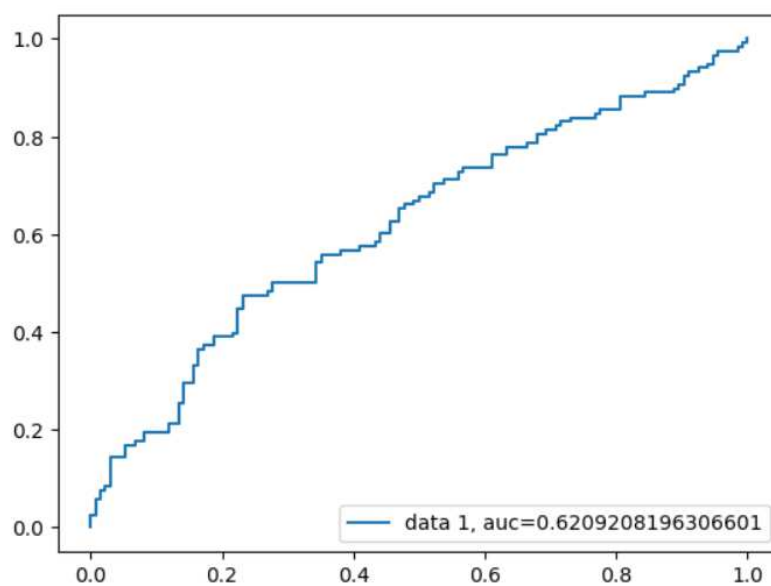
<sup>2</sup> <https://github.com/lanabaric/autism-smri>

## 4. Rezultati

Za skup značajki koji se sastojao od svih anatomskih značajki koje su izvučene iz sMRI-a, izlaz klasifikacijskog modela dao je 87 i 65 točnih predikcija te 47 i 53 pogrešnih (Slika 4.1). Točnost je iznosila 60%, preciznost za sudionike bez autizma 62%, a za one s autizmom 58%. Ukoliko u setu za testiranje ima sudionika koji imaju autizam, ovaj model će ih prepoznati u 55% slučajeva. ROC-AUC rezultat iznosi 0.6209 (Slika 4.2).

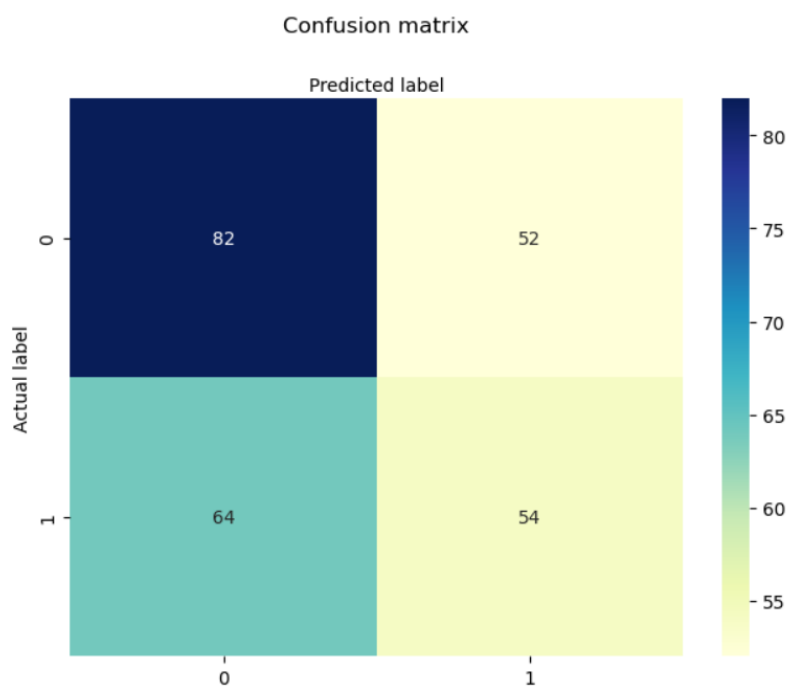


Slika 4.1 matrica zabune za sve anatomske značajke

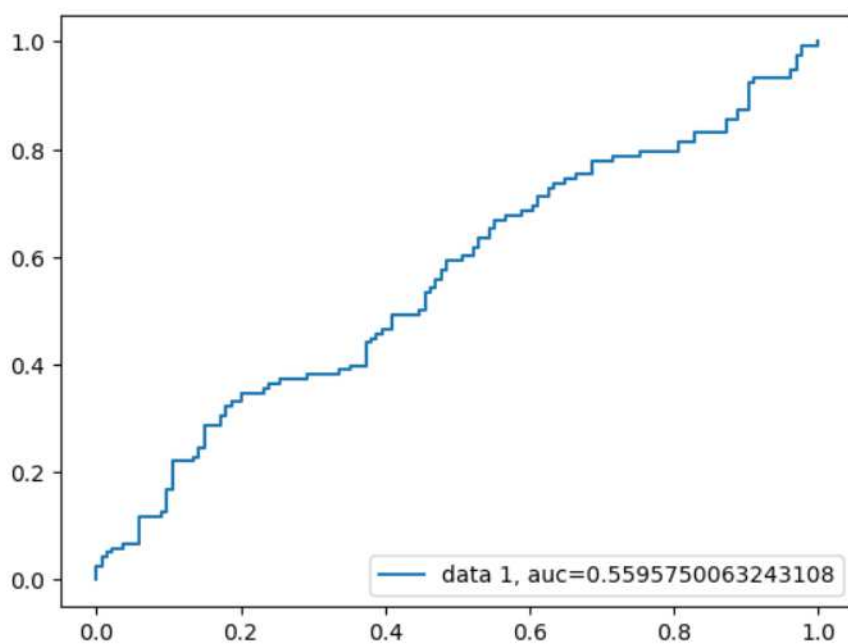


Slika 4.2 ROC krivulja za sve anatomske značajke

Za skup značajki koje se odnose na kortikalnu debljinu i površinu mozga, izlaz klasifikacijskog modela dao je 82 i 54 točnih te 52 i 64 pogrešnih predikcija (Slika 4.3). Točnost je iznosila 54%, preciznost za sudionike bez autizma 56%, a za one s autizmom 51%. Ako u setu za testiranje ima sudionika s autizmom, model će ih prepoznati u 48% slučajeva. ROC-AUC rezultat iznosi 0.5596 (Slika 4.4).



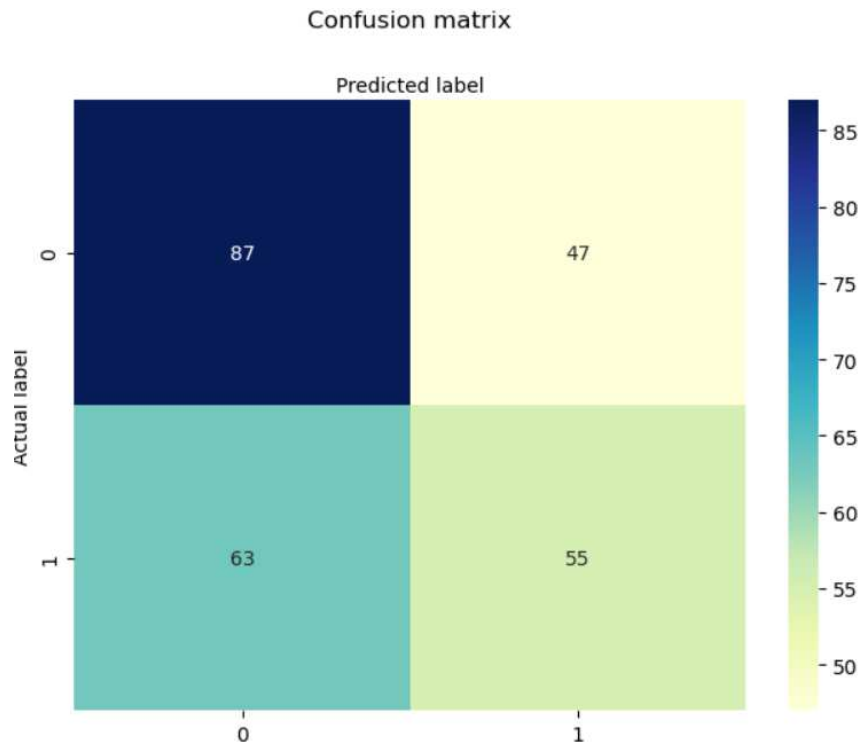
Slika 4.3 matrica zabune za značajke povezane s kortikalnom debljinom i površinom



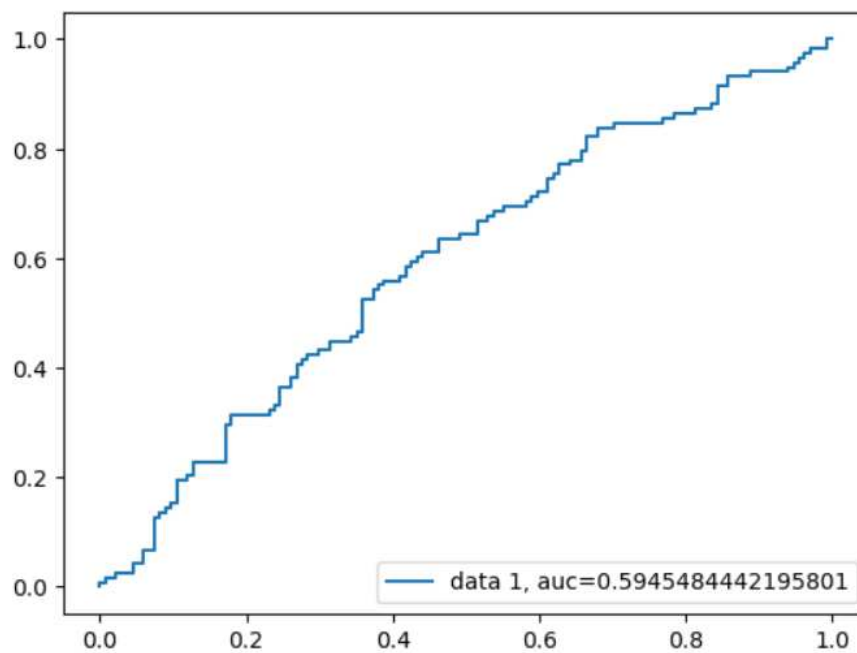
Slika 4.4 ROC krivulja za značajke povezane s kortikalnom debljinom i površinom



Za one značajke koje se odnose na volumen mozga, izlaz klasifikacijskog modela dao je 87 i 55 točnih te 47 i 63 pogrešnih predikcija (Slika 4.5). Točnost je iznosila 56%, preciznost za sudionike s autizmom 54% dok za one bez 58%. Ukoliko se u setu za testiranje nalaze sudionici koji imaju autizam, oni će biti otkriveni u 47% slučajeva. ROC-AUC rezultat iznosi 0.5945 (Slika 4.6).

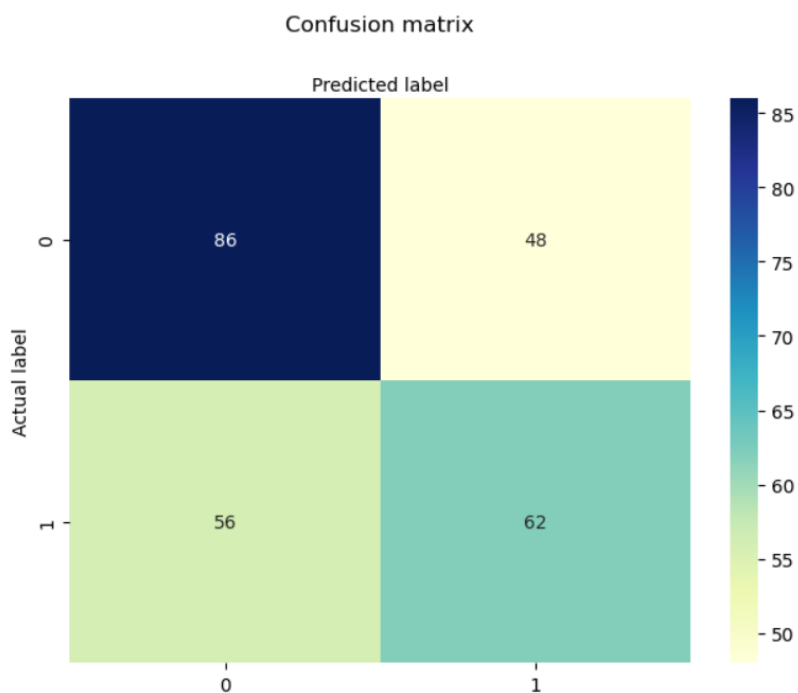


Slika 4.5 matrica zabune za značajke povezane s volumenom mozga

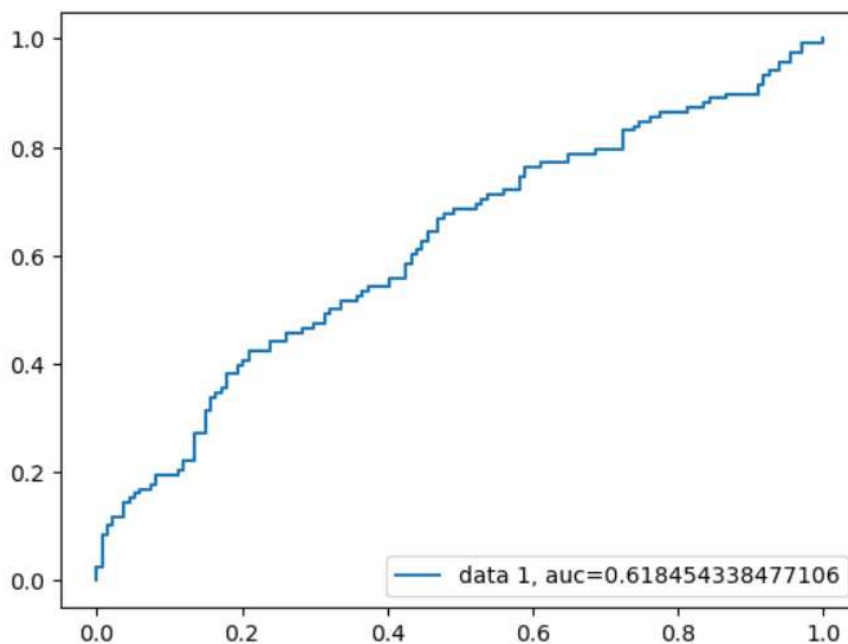


Slika 4.6 ROC krivulja za značajke povezane s volumenom mozga

Kada su u značajke uključeni i svi demografski podaci sudionika (spol, godine i lokacija), izlaz klasifikacijskog modela dao je 86 i 62 točne predikcije te 48 i 56 pogrešnih (Slika 4.7). Točnost je iznosila 59%, preciznost za sudionike s autizmom 56%, a za one bez 61%. Ako se u setu za testiranje nalaze sudionici s autizmom, oni će biti prepoznati u 53% slučajeva. ROC-AUC rezultat iznosi 0.6185 (Slika 4.8).

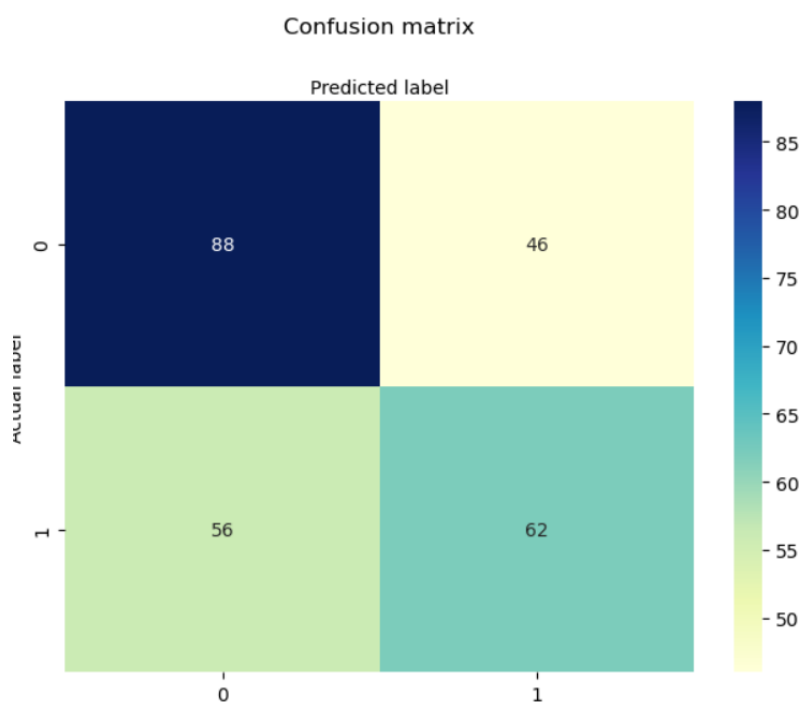


Slika 4.7 matrica zabune za demografske podatke i anatomske značajke

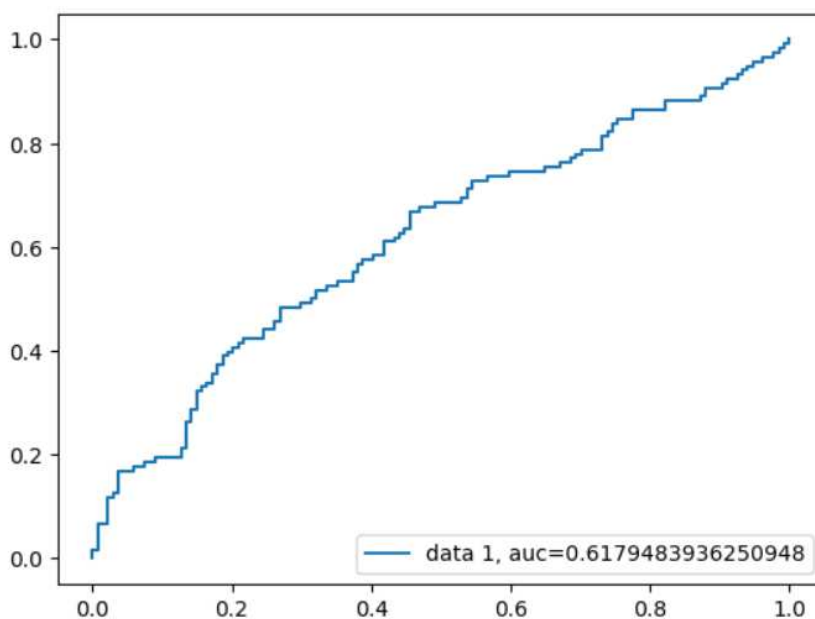


Slika 4.8 ROC krivulja za demografske podatke i anatomske značajke

Za skup značajki koje su se sastojale od godina sudionika te anatomskih značajki mozga, klasifikacijski model dao je 88 i 62 točne predikcije te 46 i 56 pogrešnih (Slika 4.9). Točnost je iznosila 60%, preciznost za sudionike s autizmom 57%, dok za one bez 61%. Ako se u skupu za testiranje nalaze sudionici s autizmom, oni će biti prepoznati u 53% slučajeva. ROC-AUC rezultat iznosi 0,6179 (Slika 4.10).

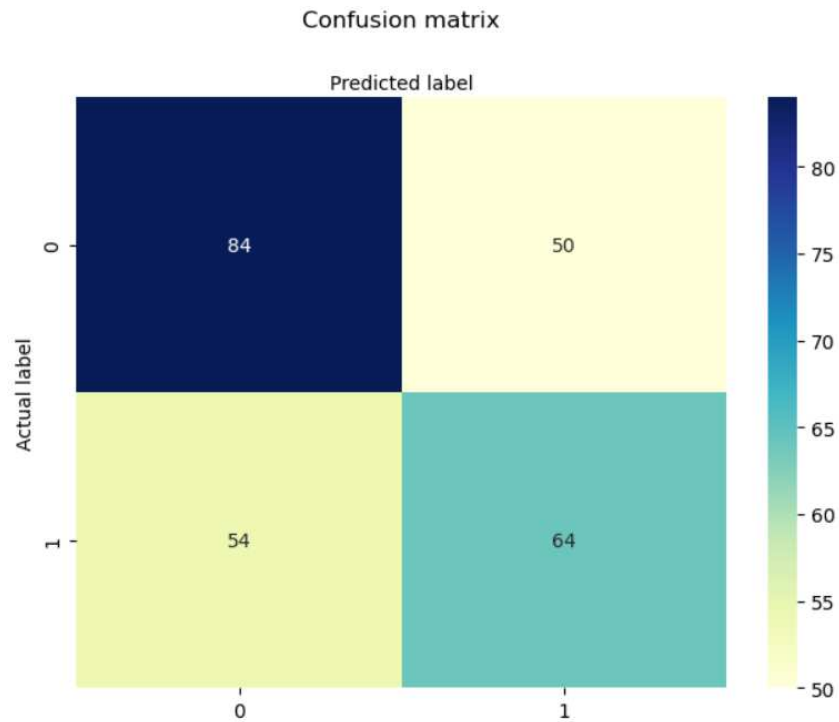


Slika 4.9 matrica zabune za godine sudionika i anatomske značajke

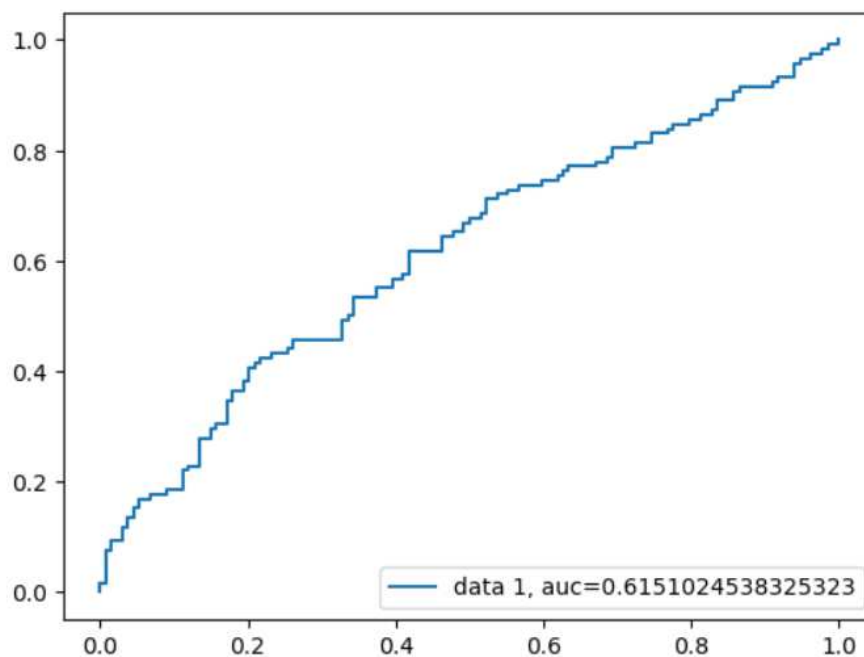


Slika 4.10 ROC krivulja za godine sudionika i anatomske značajke

Za skup značajki koji se sastoji od spola sudionika te anatomskih značajki mozga, klasifikacijski model dao je 84 i 64 točnih predikcija te 50 i 54 pogrešnih (Slika 4.11). Točnost je 59%, preciznost za sudionike s autizmom 56%, a za one bez 61%. Ukoliko se u skupu za testiranje nalaze osobe s autizmom, one će biti otkrivene u 54% slučajeva. ROC-AUC rezultat iznosi 0,6151 (Slika 4.12).



Slika 4.11 matrica zabune za spol sudionika i anatomske značajke



Slika 4.12 ROC krivulja za spol sudionika i anatomske značajke

## 4.1. Usporedba rezultata

Dobiveni rezultati nisu baš dobri – najveća točnost iznosi tek 62%, a i preciznosti se kreću između 51% i 62%. Isto tako, dobiveni AUC rezultati nisu visoki i bliže su 0.5, što sugerira slučajnost, nego 1,0, što označava savršen model. Unatoč tome što su rezultati lošiji, oni se ne razlikuju previše od onih dobivenih u članku Traut i sur. [3]. Tamo su došli do zaključka da funkcijski MRI više pridonosi predikciji autizma nego anatomske. Njihov AUC rezultat pri korištenju modela predikcije, koji je nastao kombiniranjem 10 najboljih rješenja njihovog izazova, sa samo anatomske značajkama iznosi 0.66, što se ne razlikuje jako puno od dobivenih 0,62. Također, autori članka su došli do zaključka da uključivanje demografskih podataka ima jako mali utjecaj na rezultat predikcije što je vidljivo i u dobivenim rezultatima.

## Zaključak

Unatoč tomu što dobiveni rezultati nisu idealni, može se zaključiti da je MRI bitan i koristan izvor za proučavanje tematike autizma. Najbolji rezultati su dobiveni kada su se kao značajke u modelu strojnog učenja temeljenom na logističkoj regresiji koristile sve dostupne anatomske značajke mozga. Tada je točnost bila 60%, no to je i dalje dosta manje od 80%, što bi bila granica prihvatljivosti dobre točnosti. Taj skup značajki ujedno je dao i najbolji AUC rezultat, 0,62, koji je nešto manji od dostupnog rezultata u Traut i sur. [3] koji je 0,66. Razlog tome može biti što su u [3] koristili model predikcije koji je dobiven kombinacijom 10 najboljih rješenja njihovog izazova. Razdvajanje anatomske značajke na one povezane s volumenom mozga te one povezane s kortikalnom debljinom i površinom mozga, nije pridonijelo poboljšanju predikcije modela – naprotiv, točnost i preciznost su se smanjile u oba slučaja. Dodavanjem demografskih podataka sudionika u model za predikciju, točnost i preciznost se nisu puno mijenjale, obje vrijednosti su se nalazile u rasponu od 56 do 61%. U budućim istraživanjima, bilo bi zanimljivo vidjeti kako bi dodatak podataka dobivenih funkcijskim MR slikama utjecao na ovaj model predikcije, a isto tako bi vrijedilo isprobati i neke druge algoritme klasifikacije koji bi možda ostvarili bolje rezultate. Također, bilo bi dobro koristiti veći broj sudionika no zbog slabije dostupnosti podataka, to u ovom radu nije bilo moguće.

# Literatura

- [1] Leksikografski zavod Miroslav Krleža, *autizam*, Hrvatska enciklopedija, mrežno izdanje., (2013.–2024.), Poveznica: <https://www.enciklopedija.hr/clanak/autizam> ; Pristupljeno 27.kolovoza.2024.
- [2] Jenkison, M., Chapell M. *Introduction to Neuroimaging Analysis*. 1.izdanje. Oxford University Press, 2018.
- [3] Traut, N., Heuer, K., Lemaitre, G., Beggiato, A., Germanaud, D., Elmaleh, M., Bethegnies, A., Bonnasse-Gahot, L., Cai, W., Chambon, S., Cliquet, F., Ghriss, A., Guigui, N., de Pierrefeu, A., Wang, M., Zantedeschi, V., Boucaud, A., van den Bossche, J., Kegl, B., Delorme, R., Varoquaux, G., *Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery*, NeuroImage, Volume 255, (2022, srpanj), 119171, Poveznica: <https://www.sciencedirect.com/science/article/pii/S1053811922002981#abs0001>
- [4] *Prediction*, H20.ai, Wiki, Poveznica: <https://h2o.ai/wiki/prediction/> ; pristupljeno 25.kolovoza 2024.
- [5] Šnajder, J., *Strojno učenje: 6. Logistička regresija*, UNIZG FER, ak. god. 2020./2021.
- [6] Brownlee, J., *Logistic Regression for Machine Learning*, Machine Learning Mastery, (2023, prosinac), Poveznica: <https://machinelearningmastery.com/logistic-regression-for-machine-learning/> ; pristupljeno 26.kolovoza 2024.

# Sažetak

Naslov: Predikcija poremećaja iz spektra autizma korištenjem značajki strukturnih slika magnetske rezonancije mozga

Sažetak: Autizam je kompleksan neurobiološki poremećaj ljudskog mozga koji je okarakteriziran različitim oblicima oštećenja socijalne interakcije, komunikacije i ponašanja. Kako bi osobe s autizmom dobile pravovremenu pomoć, rana dijagnoza je izrazito važna. Stoga se postavlja pitanje, može li se on dijagnosticirati i predvidjeti pomoću modela strojnog učenja te korištenjem slika magnetske rezonancije (MR) mozga. U ovom radu se fokusiralo na strukturne MR slike. Pomoću modela logističke regresije provelo se treniranje na skupu podataka ABIDE, a potom i testiranje modela. Korišteno je nekoliko skupova značajki, a najbolji rezultat je bio kada se radilo sa svim anatomskim značajkama mozga, ali bez korištenja demografskih podataka sudionika. Tada je točnost bila 60%, AUC rezultat 0,6209 te bi model logističke regresije prepoznao da osoba ima autizam u 55 % slučajeva. Dobiveni rezultati su u skladu s očekivanjima i pokazuju da bi daljnje istraživanje i razvoj modela strojnog učenja i MR slika mozga bili od velike koristi u otkrivanju autizma.

Ključne riječi: autizam, predikcija, strojno učenje, strukturne slike magnetske rezonancije mozga



# Summary

Title: Prediction of autism spectrum disorders using features from structural brain magnetic resonance imaging.

Summary: Autism is a complex neurobiological disorder of the human brain characterized by various forms of impairment in social interaction, communication, and behaviour. For individuals with autism, early diagnosis is extremely important. Therefore, the question arises whether machine learning models can be used to diagnose and predict autism with the help of brain magnetic resonance (MR) imaging. This study focused on structural MR images. A logistic regression model was trained on the ABIDE dataset and then tested. Several feature sets were used, and the best result was achieved when working with all anatomical brain features, but without using the demographic data of the participants. In that case, accuracy was 60%, the AUC score was 0.6209, and the logistic regression model would recognize that a person has autism in 55% of cases. The results obtained are in line with expectations and indicate that further research and development of machine learning models and brain MR images would be of great benefit in detection of autism.

Keywords: autism, machine learning, prediction, structural MR images

## Skraćenice

MRI *Magnetic Resonance Imaging*  
rezonancije

oslikavanje magnetske

sMRI *structural magnetic resonance imaging*  
magnetske rezonancije

strukturno oslikavanje

ASD *Autism Spectrum Disorder*

poremećaj iz spektra autizma