

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1270

**PRIMJENA STROJNOG I DUBOKOG UČENJA U
KLASIFIKACIJI BIPOLARNOG POREMEĆAJA,
SHIZOFRENIJE I ZDRAVIH ISPITANIKA NA TEMELJU
PODATAKA MAGNETSKE REZONANCIJE MOZGA**

Lana Bartolović

Zagreb, lipanj 2026.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1270

**PRIMJENA STROJNOG I DUBOKOG UČENJA U
KLASIFIKACIJI BIPOLARNOG POREMEĆAJA,
SHIZOFRENIJE I ZDRAVIH ISPITANIKA NA TEMELJU
PODATAKA MAGNETSKE REZONANCIJE MOZGA**

Lana Bartolović

Zagreb, lipanj 2026.

DIPLOMSKI ZADATAK br. 1270

Pristupnica: **Lana Bartolović (0036538315)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: prof. dr. sc. Alan Jović

Zadatak: **Primjena strojnog i dubokog učenja u klasifikaciji bipolarnog poremećaja, shizofrenije i zdravih ispitanika na temelju podataka magnetske rezonancije mozga**

Opis zadatka:

Bipolarni poremećaj i shizofrenija složeni su psihijatrijski poremećaji s djelomično preklapajućim kliničkim obilježjima, što otežava njihovo razlikovanje prilikom dijagnosticiranja. Unatoč napretku u neuroznanosti i kliničkoj praksi, dijagnostika se i dalje u velikoj mjeri temelji na kliničkoj procjeni simptoma, bez jasnih bioloških markera koji bi omogućili objektivno razlikovanje ovih poremećaja. Razvoj metoda analize podataka magnetske rezonancije mozga i primjena suvremenih računalnih pristupa otvaraju mogućnost preciznije i objektivnije potpore dijagnostičkom procesu. U ovom diplomskom radu potrebno je istražiti mogućnost automatskog razlikovanja ispitanika s bipolarnim poremećajem i shizofrenijom od zdravih ispitanika na temelju podataka magnetske rezonancije mozga. Razmatrat će se različite vrste strukturnih značajki mozga te primijeniti metode redukcije dimenzionalnosti i odabira značajki. Primijenit će se i usporediti više metoda strojnog učenja, uključujući binarne i višeklasne klasifikatore. Modeli se trebaju temeljiti na pojedinačnim i kombiniranim skupovima značajki. Također se treba ispitati mogućnosti primjene odgovarajućih metoda dubokog učenja (npr. konvolucijske neuronske mreže). Budući da je u medicinskom kontekstu važna interpretabilnost modela, u radu će se istražiti i metode objašnjavanja modela, poput analize važnosti značajki te tehnika vizualizacije aktivacija (npr. Grad-CAM) na slikovnim podacima, kako bi se identificirale moždane regije koje najviše doprinose odluci modela. Rezultate izgrađenih modela potrebno je usporediti s rezultatima iz srodne literature.

Rok za predaju rada: 26. lipnja 2026.

*Zahvaljujem svom mentoru na stručnom vodstvu i pomoći tijekom izrade ovog
diplomskog rada.*

*Također zahvaljujem svojoj obitelji na kontinuiranoj podršci i razumijevanju tijekom
studija.*

*Svojim prijateljicama iz klupe zahvaljujem na divnih pet godina i svim iskustvima koje
smo prošle zajedno. Hvala vam što ste uljepšale ovaj važan dio mog života.*

Sažetak

Primjena strojnog i dubokog učenja u klasifikaciji bipolarnog poremećaja, shizofrenije i zdravih ispitanika na temelju podataka magnetske rezonancije mozga

Lana Bartolović

Primjena metoda strojnog i dubokog učenja značajno bi olakšala svakodnevni život i liječnicima i pacijentima. Ovaj se rad bavi istraživanjem razlikovanja zdravih, bipolarnih i shizofrenih osoba koristeći podatke s magnetske rezonancije (MR) mozga. Klasično strojno učenje koristilo je brojčane značajke iz MR snimaka, a duboko učenje koristilo se VBM mapama sive i bijele tvari (slikovni podaci). Rezultati su pokazali da je najlakše razlikovati shizofrene osobe od zdravih, s točnošću do ~74%, dok su ostali zadaci puno teži s točnošću nerijetko blago iznad razine slučajnog pogađanja. Isprobani modeli dubokog učenja nisu rezultirali značajnim poboljšanjem u odnosu na klasično strojno učenje. Grad-CAM analiza objašnjivosti modela pokazala je da se neki modeli fokusiraju na nevažna područja pri klasifikaciji. Sveukupno, rezultati ukazuju na potencijal navedenih metoda, ali i naglašavaju trenutna ograničenja.

Ključne riječi: bipolarni poremećaj; shizofrenija; neuroimaging; strojno učenje; duboko učenje; redukcija dimenzionalnosti; XAI; Grad-CAM; višeklasna klasifikacija

Abstract

Application of machine learning and deep learning in the classification of bipolar disorder, schizophrenia, and healthy subjects based on brain magnetic resonance imaging data

Lana Bartolović

The application of machine learning and deep learning methods could significantly facilitate the daily lives of both physicians and patients. This thesis investigates the differentiation between healthy individuals and individuals with bipolar disorder or schizophrenia using brain magnetic resonance imaging (MRI) data. Classical machine learning methods used numerical features extracted from MRI scans, while deep learning methods were applied to voxel-based morphometry (VBM) maps of gray and white matter (image data). The results showed that individuals with schizophrenia were the easiest to distinguish from healthy individuals, with an accuracy of up to approximately 74%, while the other classification tasks were considerably more challenging, often achieving accuracies only slightly above the level of random guessing. The tested deep learning models did not yield significant improvements over classical machine learning methods. Grad-CAM analysis of model explainability showed that some models focused on irrelevant regions during classification. Overall, the results indicate the potential of these methods while also highlighting their current limitations.

Keywords: bipolar disorder; schizophrenia; neuroimaging; machine learning; deep learning; dimensionality reduction; XAI; Grad-CAM; multiclass classification

Sadržaj

Sažetak	1
Abstract	2
1. Uvod	5
2. Bipolarni poremećaj i shizofrenija	7
2.1. Bipolarni poremećaj	7
2.2. Shizofrenija	9
2.3. Usporedba	11
3. Magnetska rezonancija mozga	13
4. Metode strojnog i dubokog učenja	15
4.1. Strojno učenje	15
4.1.1. Nadzirano strojno učenje	15
4.1.2. Nenadzirano strojno učenje	16
4.1.3. Korišteni algoritmi strojnog učenja	16
4.2. Duboko učenje	22
4.2.1. Umjetne neuronske mreže (ANN)	22
4.2.2. Konvolucijske neuronske mreže (CNN)	23
4.2.3. Učenje ANN-ova	25
4.2.4. Objašnjivost	27
5. Primjena umjetne inteligencije u neuropsihijatriji	30
6. Skup podataka	35

7. Implementacija	37
7.1. Klasično strojno učenje	37
7.2. Duboko učenje i objašnjivost	40
7.2.1. Učenje od početka	42
7.2.2. Prednaučene mreže	44
7.2.3. Objašnjivost	48
8. Rezultati i rasprava	50
8.1. Klasično strojno učenje	50
8.2. Duboko učenje	52
8.2.1. Učenje od početka	52
8.2.2. Prednaučene mreže	56
8.3. Usporedba s postojećim istraživanjima	61
9. Zaključak	64
Literatura	66
Izjava o korištenju umjetne inteligencije	71

1. Uvod

Bipolarni poremećaj i shizofrenija ubrajaju se među teže psihijatrijske poremećaje s velikim utjecajem na kvalitetu života oboljelih. Bipolarni poremećaj karakteriziraju izmjenjujuća stanja manije i depresije, dok shizofreniju karakteriziraju simptomi poput psihoze, zablude i poremećaji mišljenja. Neki od pogođenih mogu imati i druge mentalne poremećaje, kao i dodatne simptome, zbog čega su dijagnoza i adekvatno liječenje nužni. Problem predstavlja činjenica da je oba poremećaja teško dijagnosticirati – čak 69% pacijenata s bipolarnim poremećajem je inicijalno krivo dijagnosticirano, a trećina ih dobije ispravnu dijagnozu tek nakon 10 i više godina [1]. Slično je i kod shizofrenije, gdje se inicijalna dijagnoza promijeni kod 36–51% pacijenata [2]. To uzrokuje da se pacijente krene liječiti za druge bolesti (najčešće je to depresija), što može značajno pogoršati njihovo stanje.

Posebno je izazovno razlikovanje bipolarnog poremećaja od shizofrenije, budući da oba poremećaja mogu imati slične simptome, osobito u ranim fazama bolesti. Jedno od mogućih rješenja bila bi primjena metoda strojnog i dubokog učenja u medicini, točnije u detekciji poremećaja. Algoritam koji bi naučio detektirati takve poremećaje imao bi široku primjenu, budući da bi uštedio vrijeme liječnicima, ali još važnije i pacijentima koji više ne bi trošili godine na pogrešne terapije i pokušaje liječenja bolesti koje nemaju, dok im simptomi postaju sve lošiji.

U ovom radu istražuje se primjena metoda strojnog i dubokog učenja za razlikovanje zdravih kontrolnih ispitanika, osoba s bipolarnim poremećajem i osoba sa shizofrenijom pomoću podataka dobivenih iz MR slika mozga. Uspoređuju se klasični pristupi strojnog učenja temeljeni na kombinacijama značajki s pristupima dubokog učenja koji koriste konvolucijske neuronske mreže, uključujući i modele prednaučene na velikim skupovima medicinskih slika. Kako bi se model učinio klinički primjenjivijim, upotri-

jebljena je metoda objašnjenja modela Grad-CAM koja omogućuje vizualizaciju područja mozga na koja se model najviše oslanja prilikom donošenja odluke. U nastavku je dan kratki uvod u bipolarni poremećaj i shizofreniju, magnetsku rezonanciju mozga te metode strojnog i dubokog učenja u medicini, nakon čega slijedi analiza znanstvenih radova koji se bave sličnom temom. Na kraju je detaljno objašnjena implementacija primijenjenih metoda u ovom istraživanju.

2. Bipolarni poremećaj i shizofrenija

2.1. Bipolarni poremećaj

Bipolarni afektivni poremećaj (BAP; ranije poznat kao manična depresija) je kronični poremećaj raspoloženja koji karakteriziraju česte promjene raspoloženja. Kod osoba koje boluju od bipolarnog poremećaja izmjenjuju se stanja depresije i stanja snažno povišenog raspoloženja – ako stanje priliči psihozi, naziva se manija (engl. *mania*), dok ako ne utječe značajno na ponašanje naziva se hipomanija (engl. *hypomania*). Od bipolarnog poremećaja boluje 1-5% svjetske populacije [3].

Bipolarni poremećaj prvi put spominje Hipokrat u antičkoj Grčkoj, koji je prepoznao dva raspoloženja: melankoliju (depresiju) i maniju (mješavinu ljutnje i euforije). On je prvi osvijestio da su mentalni poremećaji biološki, odnosno da se događaju unutar tijela – smatrao je da su uzrokovani djelovanjem tjelesnih sokova (engl. *humors*) – krvi, sluzi, žute i crne žuči na mozak. Po njemu, žuta žuč potiče stanje manije, a crna melankoliju. Uz njega, Aretaeus iz Cappadocie je istaknuo postojanje spektra raspoloženja, gdje su melankolija i manija dva ekstrema te je opisivao prijelaze između njih; također kako su osobe nakon melankolije sklone ući u stanje manije. Obojica su po prvi put povezala ta dva stanja jedno s drugim, kao i s uzrokom – kemijska neuravnoteženost u mozgu, a ne vanjski utjecaji.

Trenutno se vjeruje da ne postoji jedan uzrok bipolarnog poremećaja, već da se radi o više manjih faktora u međudjelovanju. Od njih je 70–90% genetski [4], a vanjski utjecaji koji doprinose mogućnosti razvoja bipolarnog poremećaja su zlostavljanje u djetinjstvu i dugoročna izloženost stresu. Postoje sličnosti između bipolarnog poremećaja i depresije što se tiče genetskih uzroka: kod oba poremećaja radi se o jakoj nasljednosti; rizik da će

osoba oboliti od jednog od njih je značajno viša ako je osoba obiteljski povezana s nekim tko također ima dotični poremećaj. Nadalje, jedan te isti gen/mutacija mogu povećati rizik za oba poremećaja, npr. SNP (engl. *single nucleotide polymorphism*, promjena jednog nukleotida) na genu CACNA1C je povezana i s bipolarnim poremećajem i depresijom.

Postoji više tipova bipolarnog poremećaja:

- Bipolarni poremećaj tip 1 (engl. *bipolar I disorder*)

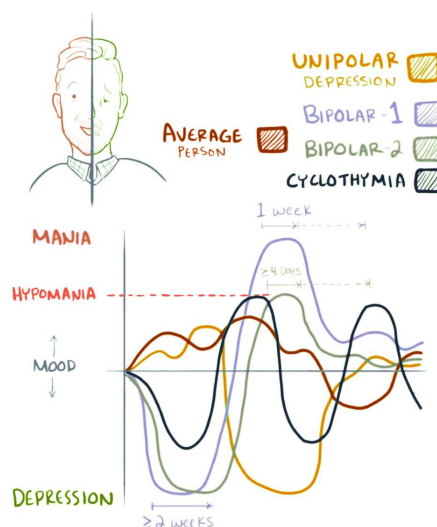
Radi se o poremećaju s barem jednom maničnom epizodom koja traje barem 7 dana. Epizode depresije su uglavnom prisutne, ali nisu nužne za dijagnozu. Većina istraživanja o bipolarnom poremećaju proučava ovaj tip.

- Bipolarni poremećaj tip 2 (engl. *bipolar II disorder*)

Kod tipa 2 umjesto manije događa se hipomanija i barem jedna depresivna epizoda. Stanje hipomanije obično ne uzrokuje ozbiljnu socijalnu štetu te nije praćeno psihozom. To dovodi do toga da je ovaj tip puno teže dijagnosticirati jer su stanja hipomanije manje zamjetna u odnosu na jaču depresiju.

- Ciklotimija (engl. *cyclothymia*)

Karakteriziraju ju stanja hipomanije i slabe depresije, tj. ovdje nema velike depresivne epizode (engl. *major depressive episode*).



Slika 2.1. Osmosis,2016. Preuzeto s Wikimedia Commons [5]

Slika 2.1. dobro prikazuje odnos tipova bipolarnog poremećaja, depresije i prosječne osobe: bipolarni poremećaj ima snažne razlike između stanja, dok su kod prosječne osobe stanja puno ujednačenija. Za razliku od njih, kod depresije prevladava jedno stanje.

Glavna metoda liječenja osoba s bipolarnim poremećajem je propisivanje stabilizatora raspoloženja, poput litija, valproata i lamotrigina. Litij je najčešći i smatra se učinkovitim kod liječenja većine maničnih stanja. Postoje ideje zašto litij ne djeluje kod svih pacijenata: istraživanja su pokazala da varijante već spomenutog gena CACNA1C mogu utjecati na učinkovitost litija u regulaciji cirkadijalnih ritmova: u stanicama s C alelom litij ne pojačava amplitudu ritmova kao u kontrolnim stanicama [6].

Stabilizatori se koriste za dugoročnu terapiju, no nisu najbolje rješenje za depresivna stanja kod bipolarnih osoba. Također to nisu ni antidepresivi jer oni mogu potaknuti da osoba prijeđe iz depresivnog stanja u (hipo)manično. Umjesto njih koriste se atipični antipsihotici. Osobe s bipolarnim poremećajem imaju povišen rizik od suicida; čak 60 puta veći postotak samoubojstava događa se kod bipolarnih osoba u odnosu na cijelu populaciju [7]. Uz to, mnoge bipolarne osobe također boluju od drugih psihičkih poremećaja, poput anksioznosti, opsesivno-kompulzivnog poremećaja, poremećaja ličnosti, ADHD-a, a mnogi postaju ovisnici u pokušaju liječenja vlastitim metodama. Osim psihičkih, često su prisutni i tjelesni problemi, uključujući dijabetes, kardiovaskularne bolesti i bolesti štitnjače.

2.2. Shizofrenija

Shizofrenija je kronični mentalni poremećaj koji primarno pogađa sposobnost razmišljanja. Oboljeli se susreću s otežanom percepcijom stvarnosti, što uzrokuje promjene ponašanja i neugodne osjećaje, kao i slabo izražavanje emocija. Od shizofrenije boluje 0.3-0.7% svjetske populacije [8].

Za razliku od bipolarnog poremećaja, shizofrenija se kreće spominjati tek početkom 19. stoljeća, no samo kao uranjena demencija. Kad je postalo jasno da se radi o sasvim drugom poremećaju, Eugen Bleuler ga je 1908. godine nazvao shizofrenijom. Etimološki naziv opravdava simptome, budući da dolazi od grčkih *schizein*, odnosno podjela i

phren, odnosno um, mislivši se pritom na podijeljenost između razmišljanja, percepcije i osobnosti. Ovisno o prisutnim simptomima, shizofrenija se dijelila na 5 podtipova, no 2013. godine je ta podjela ukinuta budući da se simptomi značajno preklapaju i mijenjaju kroz vrijeme. Umjesto podjele na podtipove, suvremeni pristup opisuje shizofreniju kroz nekoliko skupina simptoma:

- Pozitivni simptomi

Odnose se na iskustva koja nisu prisutna u zdravih osoba, poput deluzija i halucinacija. Halucinacije primarno pogađaju osjet sluha (pričinjavanje zvukova), ali nerijetko (čak dva puta češće) dolazi do halucinacija s kombinacijom osjetila. Deluzije se odnose na čvrsta, netočna uvjerenja o sebi ili drugima. Česti primjeri su osjećaj progonstva ili osjećaj velike moći. Ovi simptomi dobro reagiraju na liječenje i postaju manje prisutni s napretkom bolesti.

- Negativni simptomi

Negativni simptomi se manifestiraju kao odsutnost normalnih emocionalnih reakcija. Oni uključuju tzv. 5A: ravnodušnost (apatija), siromaštvo govora (alogija), nemogućnost doživljavanja zadovoljstva (anhedonija), manjak želje za stvaranjem veza s ljudima (asocijalnost) te manjak motivacije (avolicija). Njih je teže tretirati i samim time više utječu na kvalitetu života. Također su prepoznati i sekundarni negativni simptomi, koji se odnose na posljedice pozitivnih simptoma, kao npr. zlouporaba opojnih sredstava.

- Kognitivni simptomi

Otprilike 70% oboljelih pri početku bolesti osjeti i kognitivne deficite, najviše povezane uz verbalnu memoriju (apstrakcije povezane s jezikom) te percepciju sebe i ljudi oko sebe (npr. interpretacija emocija kod ostalih osoba nerijetko je izazov).

Raznolikost simptoma odražava i složenost samoga poremećaja – točan uzrok i dalje nije poznat. Najprihvaćenije teorije tvrde da shizofrenija nastaje kao posljedica složene interakcije genetskih (pretjerano lučenje dopamina; abnormalnosti u NMDA receptorima) i okolišnih čimbenika (kronični stres; komplikacije pri rođenju), ali i samog razvoja mozga (manjak sive tvari; abnormalan razvoj astrocita).

Najčešći oblik liječenja uključuje primjenu antipsihotika u kombinaciji s psihosocijalnom podrškom i intervencijama. Antipsihotici djeluju primarno na neurotransmisiju dopamina i serotonina te su učinkoviti u ublažavanju pozitivnih simptoma bolesti, no negativni i kognitivni simptomi i dalje bivaju prisutni. S obzirom na to, kao i već spomenute sekundarne negativne simptome, životni vijek osoba oboljelih od shizofrenije je smanjen za čak 10 do 28 godina, a sama kvaliteta života značajno narušena. To potvrđuju statistike poput 85% nezaposlenosti kod oboljelih, više stope suicida, čak 5-6% i čestog oboljenja od kardiovaskularnih bolesti [8].

2.3. Usporedba

Ova dva poremećaja dijele određene značajke, ali se razlikuju prema dominantnim simptomima, tijeku bolesti i pristupu liječenju. Oba poremećaja imaju složenu etiologiju u kojoj važnu ulogu imaju genetski i okolišni čimbenici. Nadalje, oba su povezana s povećanim rizikom suicida, posljedično razvijenim bolestima te zahtijevaju dugotrajno liječenje.

Unatoč navedenim sličnostima, riječ je o poremećajima koji zahvaćaju različite aspekte psihičkog funkcioniranja. Bipolarni poremećaj primarno je poremećaj raspoloženja, obilježen izmjenom depresivnih i (hipo)maničnih epizoda. Nasuprot tome, shizofrenija se smatra poremećajem mišljenja i percepcije, pri čemu dominiraju halucinacije, deluzije i gubitak emocionalnog izražavanja. Kod osoba s bipolarnim poremećajem psihotični simptomi mogu biti prisutni tijekom težih maničnih ili depresivnih epizoda, ali su u pravilu povezani s promjenama raspoloženja. Kod shizofrenije psihotični simptomi nisu nužno vezani uz raspoloženje te mogu biti prisutni tijekom većeg dijela bolesti.

Razlikuju se i terapijski pristupi. Temelj liječenja bipolarnog poremećaja čine stabilizatori raspoloženja, dok se antipsihotici primjenjuju ponajprije tijekom maničnih epizoda ili kod prisutnosti psihotičnih simptoma. S druge strane, antipsihotici predstavljaju osnovu liječenja shizofrenije. Povezano s uspješnosti liječenja, osobe sa shizofrenijom u prosjeku imaju izraženije poteškoće u svakodnevnom funkcioniranju i slabiji funkcionalni oporavak nego osobe s bipolarnim poremećajem, kod kojih između pojedinih epizoda može doći do gotovo potpunog oporavka.

Posljednjih godina genomska istraživanja pokazala su da bipolarni poremećaj i shizofrenija dijele dio genetske podloge. Istraživanja su identificirala niz zajedničkih gena povezanih s oba poremećaja, uključujući gen *CACNA1C*. [6], [9]. Takvi nalazi upućuju na postojanje zajedničkih bioloških mehanizama koji doprinose razvoju oba poremećaja, unatoč njihovim različitim kliničkim manifestacijama. Upravo zbog toga bipolarni poremećaj i shizofrenija često se promatraju kao poremećaji koji se djelomično preklapaju na molekularnoj i genetskoj razini, što predstavlja važan smjer suvremenih istraživanja u psihijatrijskoj genetici.

3. Magnetska rezonancija mozga

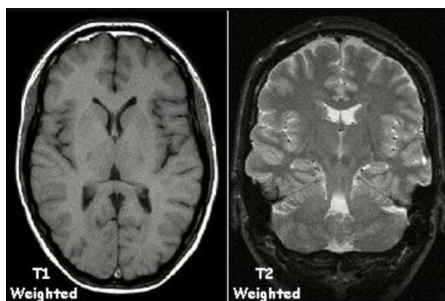
Magnetska rezonancija (engl. *Magnetic Resonance Imaging*, MRI) mozga je neinvazivna tehnika koja u medicini služi za dobivanje slika anatomije i fizioloških procesa u tijelu. Za razliku od CT skeniranja, ne koristi ionizirajuće zračenje pa se zato i preferira, no budući da je MRI dugotrajan proces koji se sastoji od ulaska u dugu i usku "cijev", kao što je prikazano na slici 3.1., koja proizvodi neugodnu buku, pacijentima s uvjetima poput klaustrofobije nije moguće napraviti MRI. Također, osoba na MRI pregledu ne smije imati metalne ili magnetske predmete na sebi, poput naušnica, satova; a u nekim slučajevima ni predmete u sebi (pacemakeri, zubni implantati). Razlog tomu je činjenica da MRI koristi snažno magnetsko polje pa metal može biti privučen ili se može krenuti zagrijavati.



Slika 3.1. MRI uređaj. Preuzeto s Technology Networks [10]

Osnovni princip MRI-ja temelji se na svojstvima atomskih jezgri, osobito jezgri vodika zato što je vodik jako prisutan u ljudskom tijelu (npr. u vodi i mastima) i njegova jezgra se sastoji samo od jednog protona. Kad se tijelo postavi unutar jakog magnetskog polja, npr. u MRI uređaj, magnetski momenti protona se poravnaju s tim poljem. Nakon toga se šalje radio frekvencijski impuls koji pobuđuje protone – oni apsorbiraju danu energiju i privremeno pređu u stanje više energije, a pri povratku u ravnotežno stanje emitiraju signal koji se koristi za konstrukciju slike. Različitom tkivu je potrebna drukčija količina vremena da se vrati u ravnotežno stanje; to se iskazuje T1 i T2 vremenima

relaksacije. Ovisno o tim vremenima, mogu se dobiti različite slike – T1-mjerene slike jasno prikazuju anatomiju mozga, mekog i masnog tkiva, dok se T2-mjerene slike više koriste u patologiji, npr. za detekciju tumora: na njima je voda jako svijetla pa se tumori (koji sadrže više vode od njima okolnog tkiva) bolje vide.



Slika 3.2. Primjer T1 i T2 mjerenih MR slika ljudskog mozga. Prilagođeno iz izvornika u javnom vlasništvu [11]

Na slici 3.2. se vidi razlika između T1 i T2 mjerenih slika – na T1-mjerenim slikama (lijevo) masno tkivo i bijela tvar su svijetli dok je voda tamna. Na T2-mjerenim slikama (desno) tekućine su vrlo svijetle, bijela tvar je tamna (tamnija od sive tvari).

4. Metode strojnog i dubokog učenja

4.1. Strojno učenje

Strojno učenje je dio umjetne inteligencije koje se bavi razvojem i učenjem statističkih algoritama koji uče na podacima te to primjenjuju na novim zadacima, odnosno mogu obavljati zadatke bez eksplicitnih uputa [12]. Drugim riječima, strojno učenje koristi se za ekstrakciju informacija iz podataka i njihovu pohranu tako da se kasnije mogu koristiti [13]. Algoritam strojnog učenja, kad se primijeni na ulazne podatke, rezultira naučenim modelom. Model je na početku definiran do na neku razinu, a uči tako da optimira funkciju gubitka na temelju podataka. Postoji više pristupa strojnom učenju: nadzirano, nenadzirano i podržano učenje. U ovom radu korišteni su algoritmi nadziranog učenja te PCA, što spada u tehniku nenadziranog učenja.

4.1.1. Nadzirano strojno učenje

Naziv nadzirano učenje odnosi se na podatke na kojima model uči. Kod nadziranog učenja, podaci su grupirani kao parovi $(x, y) = (\text{ulaz}, \text{labela})$ gdje je labela očekivani izlaz modela za dani ulaz. Zadaća modela je naći preslikavanje

$$\hat{y} = f(x) \tag{4.1}$$

Primjer je model koji uči prepoznati imenice; kod njega će ulazni podaci biti u obliku npr. (jabuka, 1), (suknja, 1), (gledati, 0), gdje 1 predstavlja klasu "imenica", a 0 predstavlja klasu "nije imenica". Prethodno opisan primjer spadao bi u problem klasifikacije, gdje se ulazne primjere pokušava svrstati u prethodno definirane klase. Ako postoje dvije klase, govorimo o binarnoj klasifikaciji, u suprotnom o višeklasnoj klasifikaciji. U ovom radu

spominju se oba problema (tri binarne klasifikacije i jedna višeklasna klasifikacija).

4.1.2. Nenadzirano strojno učenje

Kod nenadziranog strojnog učenja, podaci nemaju unaprijed definiranu labelu, već je na modelu da otkrije najsmisleniju podjelu. Koristi se za zadatke poput grupiranja, redukcije dimenzionalnosti i komprimiranja, pronalaska skrivenih obrazaca u podacima. Primjer je model koji pronalazi prirodne grupe među podacima na temelju sličnosti, poput algoritma k-sredina (engl. *k-means*).

4.1.3. Korišteni algoritmi strojnog učenja

Algoritam strojnog učenja definira preslikavanje ulaznih podataka u izlazne vrijednosti, pri čemu izlaz ovisi o vrsti zadatka. Primjerice, kod klasifikacije izlaz predstavlja skup klasa, kod regresije numeričke predikcije, a kod smanjenja dimenzionalnosti novu reprezentaciju podataka u prostoru manje dimenzije. Nadalje, algoritam definira i funkciju gubitka kojom se procjenjuje pogreška odabranog preslikavanja te postupak kojim se nalazi preslikavanje koje najbolje opisuje odnos između ulaznih i izlaznih podataka. U nastavku su opisani korišteni algoritmi strojnog učenja.

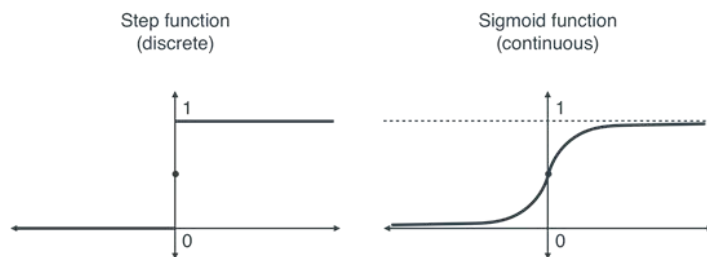
Logistička regresija

Logistička regresija jedan je od najjednostavnijih algoritama za binarnu klasifikaciju – modelira pripadnosti promatranog primjera jednoj od dvije klase. Zbog svoje jednostavnosti, često se koristi kao osnovni model za usporedbu s naprednijim metodama strojnog učenja.

Osnovna ideja logističke regresije jest izračunati linearnu kombinaciju ulaznih značajki, a zatim njezinu vrijednost preslikati u interval (0, 1) primjenom logističke funkcije, odnosno sigmoide, što je prikazano izrazom 4.2

$$\hat{y} = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}}. \quad (4.2)$$

gdje $\mathbf{w}$ predstavlja vektor težina, a b skalarni pomak. Razlog korištenja sigmoide kao aktivacijske funkcije jest njezina derivabilnost, vjerojatnosna interpretacija i sličnost funkciji praga. Na slici 4.1. prikazana je upravo usporedba funkcije praga i sigmoide, gdje se



Slika 4.1. Usporedba funkcije praga i sigmoide. Preuzeto s [14]

vidi da iako sigmoida ne daje strogu odluku kao funkcija praga, njezin oblik je dovoljno sličan, a omogućava proces učenja poznatim metodama. Nakon djelovanja sigmoide, primjer se klasificira ovisno o unaprijed definiranom pragu (najčešće 0.5) – u pozitivnu klasu ako je

$$\hat{y} \geq \text{prag}$$

Sam algoritam logističke regresije procjenjuje dva skupa parametara:

- vektor težina $\mathbf{w}$,
- skalarni pomak b ,

kroz iterativnu optimizaciju funkcije pogreške koristeći tehniku gradijentnog spusta. Korištena funkcija pogreške jest gubitak unakrsne entropije (engl. *Cross Entropy Loss*) izveden metodom maksimalne izglednosti (engl. *Maximum Likelihood Estimation*, MLE). Ideja ove metode jest pronaći vrijednosti parametara modela koje maksimiziraju vjerojatnost opaženih oznaka u dostupnom skupu podataka, odnosno koje čine opažene podatke najvjerojatnijima prema modelu. Sukladno tome, funkcija pogreške definirana je kao negativna izglednost opaženih oznaka s obzirom na skup podataka (točnije negativna log-izglednost radi lakšeg izračuna).

Budući da se radi o klasifikacijskom problemu, oznake su modelirane kao Bernoulijeve varijable, čija se vjerojatnost za jedan primjer može zapisati kao

$$P(y \mid \mu) = \mu^y (1 - \mu)^{1-y} \quad (4.3)$$

gdje je μ vjerojatnost da primjer pripada klasi 1, što zapravo odgovara izlazu iz modela:

$$P(y \mid \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b)^y (1 - \sigma(\mathbf{w}^T \mathbf{x} + b))^{1-y}, \quad (4.4)$$

A za cijeli skup primjera ukupna vjerojatnost odgovara umnošku pojedinačnih vjerojatnosti:

$$P(\mathbf{y} | \mathbf{X}) = \prod_{i=1}^N p(y^{(i)} | \mathbf{x}^{(i)}) \quad (4.5)$$

čiji se negativni logaritam 4.6 zapravo minimizira tijekom procesa učenja.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y^{(i)} \ln(\hat{y}^{(i)}) + (1 - y^{(i)}) \ln(1 - \hat{y}^{(i)})], \quad (4.6)$$

Za optimizaciju funkcije pogreške koristi se algoritam gradijentnog spusta. Gradijent funkcije pokazuje smjer njezina najvećeg rasta pa se u svrhu minimizacije funkcije pogreške parametri modela ažuriraju u smjeru suprotnom od gradijenta, sukladno izrazu 4.7

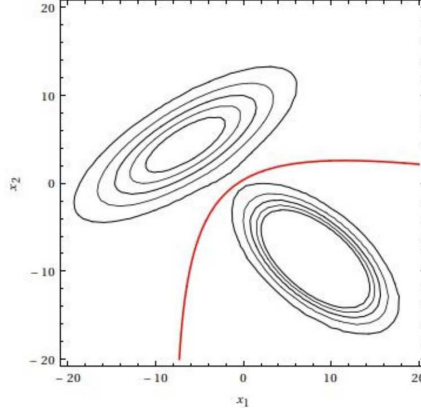
$$\nabla_{\mathbf{w}} \mathcal{L} = \frac{1}{N} \sum_{i=1}^N (\hat{y}^{(i)} - y^{(i)}) \mathbf{x}^{(i)}. \quad (4.7)$$

Kvadratna diskriminantna analiza (QDA)

QDA poput logističke regresije pripada skupini probabilističkih klasifikatora. No za razliku od logističke regresije, ona je generativan model – modelira cijelu distribuciju klasa, ne samo granicu između istih. Model pretpostavlja da su podaci unutar svake klase distribuirani prema višedimenzionalnoj normalnoj (Gaussovoj) distribuciji. Za razliku od linearne diskriminantne analize (LDA), QDA ne pretpostavlja zajedničku matricu kovarijance za sve klase, već svaka klasa ima vlastitu matricu kovarijance. Matrica kovarijance opisuje raspršenost podataka unutar pojedine klase te međusobnu povezanost njezinih obilježja. U geometrijskom smislu, matrica kovarijance određuje oblik, veličinu i orijentaciju raspodjele podataka u prostoru značajki. Time model može modelirati nelinearne granice odlučivanja između klasa, vidljivo na slici 4.2.

Za razliku od mnogih modela strojnog učenja, QDA se ne uči iterativnom optimizacijom funkcije gubitka. Umjesto toga, parametri modela procjenjuju se izravno iz podataka za učenje ranije spomenutom metodom maksimalne izglednosti. Tijekom procesa učenja za svaku klasu procjenjuju se tri skupa parametara:

- vektor srednjih vrijednosti $\boldsymbol{\mu}_k$,
- matrica kovarijance $\boldsymbol{\Sigma}_k$,



Slika 4.2. Prikaz distribucije klasa s nelinearnom granicom. Preuzeto s [15]

- apriorna vjerojatnost klase π_k .

Pretpostavi li se da uzorak $\mathbf{x}$ pripada klasi k , gustoća vjerojatnosti može se opisati višedimenzionalnom normalnom distribucijom

$$p(\mathbf{x} | k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right), \quad (4.8)$$

gdje je d broj značajki, $\boldsymbol{\mu}_k$ vektor srednjih vrijednosti klase, a Σ_k pripadna matrica kovarijance. Izraz u eksponentu predstavlja Mahalanobisovu udaljenost uzorka od središta klase te uzima u obzir međusobnu korelaciju značajki.

Kako bi se odredila najvjerojatnija klasa, primjenjuje se Bayesov poučak. Budući da je nazivnik Bayesova poučka jednak za sve klase, dovoljno je usporediti diskriminantne funkcije definirane izrazom

$$\delta_k(\mathbf{x}) = -\frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) + \ln \pi_k. \quad (4.9)$$

Pritom prvi član kažnjava klase velike raspršenosti, drugi član mjeri Mahalanobisovu udaljenost uzorka od sredine klase, dok treći član uzima u obzir apriornu vjerojatnost pojavljivanja pojedine klase. Novi uzorak klasificira se u onu klasu za koju diskriminantna funkcija poprima najveću vrijednost

$$\hat{y} = \arg \max_k \delta_k(\mathbf{x}). \quad (4.10)$$

Slučajna šuma

Stablo odlučivanja (engl. *Decision Tree*) jedan je od osnovnih algoritama strojnog učenja koji se koristi za klasifikacijske i regresijske probleme. Ideja stabla odlučivanja temelji se na uzastopnom dijeljenju skupa podataka prema vrijednostima ulaznih značajki, pri čemu se prostor značajki postupno razdvaja na manje podskupove. Svaki unutarnji čvor stabla predstavlja odluku temeljenu na određenoj značajki, grane predstavljaju ishode tih odluka, dok listovi stabla sadrže konačne predikcije modela.

Proces izgradnje stabla započinje od korijenskog čvora, gdje se odabire značajka i vrijednost praga koja najbolje razdvaja podatke prema definiranoj mjeri kvalitete podjele. U ovom istraživanju korišten je Gini indeks, koji mjeri vjerojatnost krive klasifikacije slučajno odabranog primjera iz skupa. Postupak se ponavlja u svakom čvoru dok se ne dosegne kriterij zaustavljanja.

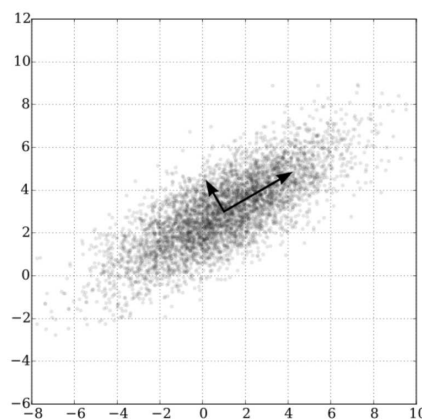
Budući da se struktura stabla snažno prilagođava dostupnom skupu za učenje, pojedinačna stabla često imaju problem velike varijance. Kako bi se smanjio utjecaj navedenog problema, koriste se metode ansambala (engl. *ensemble methods*), čija je osnovna ideja kombinirati veći broj pojedinačnih modela i njihove rezultate objediniti u jednu konačnu predikciju. Pretpostavka ovakvog pristupa jest da skup više različitih modela može ostvariti bolje performanse od pojedinačnog modela, jer pogreške pojedinih modela nisu potpuno međusobno korelirane.

Jedna od najpoznatijih metoda ansambala temeljenih na stablima odlučivanja jest slučajna šuma (engl. *Random Forest*). Slučajna šuma se sastoji od više pojedinačnih stabala učenih na različitim podskupovima primjera i značajki. Potrebno je naglasiti važnost različitosti podskupova jer se tako dobivaju raznovrsna stabla te povećava sposobnost generalizacije algoritma. Konačni izlaz slučajne šume određuje se većinskim glasovanjem svih stabala.

Analiza glavnih komponenti (PCA)

Analiza glavnih komponenti (engl. *Principal Component Analysis*, PCA) razlikuje se od prethodno spomenutih algoritama strojnog učenja jer nema za cilj provesti klasifikaciju primjera niti učiti parametre optimizacijom funkcije gubitka, već se primarno koristi za transformaciju podataka iz izvornog prostora u prostor manje dimenzionalnosti.

PCA je linearna transformacija koja nastoji pronaći novi koordinatni sustav niže dimenzije u koji se mogu preslikati podaci, pri čemu su nove osi definirane glavnim komponentama. Linearnost čini PCA računalno jednostavnom, ali sa sobom nosi i ograničenje – mogućnost pronalaska samo linearnih odnosa. Kako bi što točnije reprezentirala izvorne podatke, glavne komponente se konstruiraju linearnom kombinacijom značajki koja pokriva najveću varijancu u podacima. Svaka sljedeća glavna komponenta pruža se u smjeru najveće preostale varijance uz uvjet ortogonalnosti na svaku od već postojećih komponenti. Na slici 4.3. su vidljive upravo prve dvije glavne komponente za primjer dvodimenzionalnih podataka.



Slika 4.3. Prikaz prve dvije glavne komponente. Preuzeto s [16]

Izračun PCA-a započinje centriranjem, a po potrebi i skaliranjem značajki kako bi se uklonio utjecaj različitih mjernih skala. Nakon toga izračunava se matrica kovarijance te njezini svojstveni vektori – upravo oni definiraju smjerove glavnih komponenti. Komponente se zatim poredaju po svojstvenim vrijednostima (tj. količini objašnjene varijance), a podaci se transformiraju u prostor definiran odabranim komponentama.

U ovom istraživanju PCA se koristi kao postupak predobrade u prethodno navedenim algoritmima strojnog učenja, kao korak smanjenja broja značajki kako bi se smanjila računalna složenost i potencijalno poboljšale performanse uklanjanjem redundantnih informacija. Osim navedene primjene, PCA se često koristi i kao alat za vizualizaciju visokodimenzionalnih podataka u dvodimenzionalnom ili trodimenzionalnom prostoru, što omogućava jednostavnije uočavanje obrazaca i anomalija u podacima.

Razlog odabira ovih algoritama temelji se na potrebi usporedbe različitih pristupa klasifikaciji, od jednostavnih linearnih do složenijih nelinearnih modela. Logistička re-

gresija korištena je kao jednostavan i brz model, dok QDA omogućuje nelinearne granice odlučivanja. Slučajna šuma odabrana je zbog sposobnosti učenja složenijih odnosa među značajkama i otpornosti na šum. PCA je primijenjena radi smanjenja broja značajki s više stotina tisuća na 10, 50, 100 ili 500 (ovisno o ispitivanoj konfiguraciji; detaljnije u 7.1.), što je posebno važno u ovom slučaju zbog ograničenog broja dostupnih uzoraka.

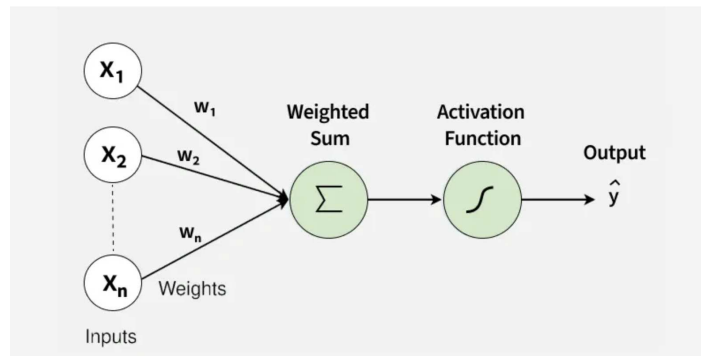
4.2. Duboko učenje

Duboko učenje je potpodručje strojnog učenja zasnovano na dubokim neuronskim mrežama. Za razliku od tradicionalnih algoritama strojnog učenja, modeli dubokog učenja mogu automatski učiti hijerarhijske reprezentacije podataka, čime se smanjuje potreba za ručnim oblikovanjem značajki. Zbog sposobnosti modeliranja složenih nelinearnih odnosa, duboko učenje danas se uspješno primjenjuje u brojnim područjima, poput računalnog vida, obrade prirodnog jezika i analize medicinskih slika.

Razvojem računalne snage i novih metoda optimizacije, umjetne neuronske mreže posljednjih su godina doživjele značajan napredak te su razvijene brojne specijalizirane arhitekture prilagođene različitim vrstama podataka i zadataka. Unatoč velikom broju postojećih varijanti, sve se one temelje na istim osnovnim principima rada umjetnih neuronskih mreža.

4.2.1. Umjetne neuronske mreže (ANN)

Umjetne neuronske mreže (engl. *Artificial neural networks*, ANN) su algoritam i način predstavljanja znanja zasnovan na ljudskom mozgu. Kako se biološki neuroni sastoje od tijela koje obavlja procesiranje te poveznica s ostalim neuronima preko kojih prima i daje informaciju, tako su organizirani i umjetni neuroni. Ukratko, svaki umjetni neuron prima informaciju u obliku težinske sume izlaza drugih slojeva od određenog broja neurona, a određenom broju neurona on daje informaciju (slika 4.4.) nakon što djeluje aktivacijskom funkcijom. Spomenuta aktivacijska funkcija mora biti nelinearna jer bi u suprotnom svaka mreža bila ekvivalentna jednom linearnom modelu te ne bi mogla rješavati složenije klasifikacijske zadatke. U praksi se koriste različite aktivacijske funkcije, poput već spomenute sigmoidne funkcije ili zglobnice (engl. *Rectified Linear Unit*,



Slika 4.4. Struktura umjetnog neurona. Preuzeto s [17]

ReLU) 4.11, pri čemu izbor funkcije ovisi o vrsti problema i arhitekturi mreže.

$$\text{ReLU}(x) = \max(0, x) \quad (4.11)$$

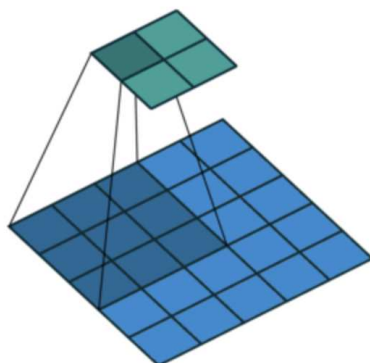
U novijim istraživanjima zglobnica istiskuje sigmoidu uglavnom jer zglobnica ne pati u tolikoj mjeri od problema nestajućeg gradijenta (engl. *vanishing gradient*), zbog čega omogućuje učenje dubljih neuronskih mreža. Osim toga, njezin izračun je jednostavniji, što smanjuje računalnu složenost i ubrzava proces učenja.

4.2.2. Konvolucijske neuronske mreže (CNN)

Klasične neuronske mreže kojima su neuroni susjednih slojeva potpuno povezani nisu praktične zbog količine veza što vodi do kompleksnog i dugog učenja. Također, potpuno povezane mreže intuitivno nisu najbolji izbor za rešetkaste podatke poput slika i videa zbog same prirode podataka. Naime, kod obrade slika prostorna udaljenost ulaza se ne smije zanemariti. Stvar je u tome da su bliži pikseli korelirani jer vrlo vjerojatno pripadaju istom objektu na slici. Kod klasičnih neuronskih mreža ne postoji način da mreža koristi tu informaciju. Zbog toga je prikladnija arhitektura mreže ona kod koje na neuron idućeg sloja utječe samo manji broj pametnije odabranih neurona ranijeg sloja.

Ovako opisane mreže nazivaju se konvolucijske neuronske mreže (engl. *Convolutional Neural Network*, CNN). Ime su dobile po operaciji konvolucije (točnije unakrsne korelacije) koju provode nad ulazom – koriste jezgru (engl. *kernel*) koja prolazi po neuronima ulaza i stvara mapu značajki (engl. *feature map*). Za 2D konvoluciju koja se koristi nad slikama, jezgra je matrica koja sadrži težine. Operacija konvolucije zapravo je težinska suma korespondentnih izlaza neurona i težina, s aktivacijskom funkcijom. Na

slici 4.5. je prikazana ideja konstruiranja novog sloja – novi neuron nastaje ovisno o manjem podskupu starih. Težine unutar jezgre jednake su za cijelu mapu značajki. Takvo

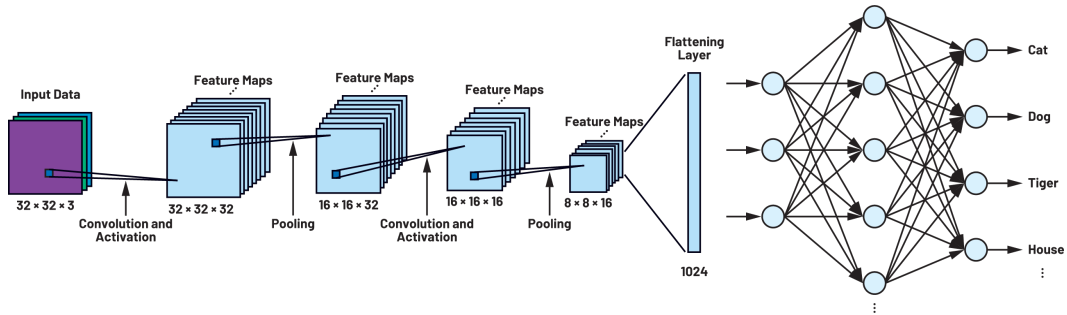


Slika 4.5. "Animation of a variation of the convolution operation" by Vincent Dumoulin and Francesco Visin, licensed under MIT. Available at: [18]

dijeljenje parametara motivirano je ponovno ljudskom biologijom (točnije ljudskim vidnim sustavom). Ljudi uspijevaju raspoznati oblike prvo detektirajući rubove, a kasnije kompleksnije kombinacije. Neuroni u vidnom korteksu reagiraju na iste vrste podražaja neovisno o njihovu položaju u vidnom polju, pa se isti mehanizam detekcije može primijeniti na cijelu sliku. Na sličan način, u konvolucijskim neuronskim mrežama ista jezgra služi za prepoznavanje iste vrste značajke bez obzira na njezinu lokaciju u slici. Pritom su konvolucijski slojevi organizirani hijerarhijski, pri čemu prvi slojevi uče prepoznavati jednostavne značajke, poput rubova i tekstura, dok kasniji slojevi njihove kombinacije koriste za prepoznavanje složenijih uzoraka, dijelova objekata i naposljetku cijelih objekata. Osim ovog intuitivnog objašnjenja, dijeljenje parametara ima i praktičnih prednosti. Ono znatno smanjuje broj parametara koje je potrebno naučiti, čime se ubrzava proces učenja. Nadalje, budući da se ista jezgra primjenjuje na velikom broju različitih položaja tijekom učenja, parametri dobivaju više podražaja za prilagodbu, što dodatno poboljšava sposobnost modela da generalizira na nove podatke.

Na slici 4.6. prikazana je tipična arhitektura konvolucijske neuronske mreže. Ulazna slika prolazi kroz niz konvolucijskih slojeva koji izdvajaju značajke, pri čemu se između njih često primjenjuju slojevi sažimanja (engl. *pooling*) kojima se smanjuju prostorne dimenzije mapa značajki. Nakon izdvajanja značajki, one se pretvaraju u jednodimenzionalni vektor pomoću sloja izravnivanja (engl. *flatten*) te se proslijeđuju potpuno povezanim slojevima za konačnu klasifikaciju. U suvremenim arhitekturama sloj flatten

često se zamjenjuje globalnim prosječnim sažimanjem (engl. *Global Average Pooling*) jer smanjuje broj parametara i nudi mreži neovisnost o dimenzijama ulaza.



Slika 4.6. Arhitektura konvolucijske neuronske mreže. Prilagođeno s [19]

4.2.3. Učenje ANN-ova

Kako bi neuronska mreža mogla nešto naučiti, potrebno je na neki način okarakterizirati njezino ponašanje, odnosno definirati koliko se njezin izlaz razlikuje od željenog. U tu se svrhu uvodi funkcija gubitka. Funkcija gubitka već je spomenuta u ranijim poglavljima, no ovdje ju je potrebno dodatno razmotriti jer ona predstavlja temelj učenja neuronske mreže. Njezina je uloga kvantificirati razliku predviđenog od očekivanog izlaza, neovisno o tome je li izlaz klasificiran točno ili netočno. Kod klasifikacijskih problema najčešće se koristi unakrsna entropija, u svom binarnom 4.6 ili višeklasnom 4.12 obliku.

$$L_{CE} = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^C y_{n,i} \log(\hat{y}_{n,i}) \quad (4.12)$$

Idealna funkcija gubitka bio bi gubitak 0-1, budući da za točne predikcije nije potrebno mijenjati model. Takav gubitak samo određuje je li predikcija točna ili netočna, a ne daje informaciju o tome koliko je predikcija netočna – jednako se kažnjavaju potpuno krive predikcije i one granične. Dodatni je problem što je gubitak 0-1 nederivabilan, pa ne pruža informaciju o tome u kojem smjeru i za koliko treba promijeniti parametre modela kako bi se gubitak smanjio. Zbog toga se pokušaj učenja uz gubitak 0-1 svodi na kombinatornu optimizaciju i iscrpno pretraživanje mogućih vrijednosti parametara, što je računalno neizvedivo. Uzevši u obzir navedena ograničenja gubitka 0-1, kao prikladna zamjena nameće se unakrsna entropija.

U prethodnom poglavlju, u kontekstu logističke regresije, već je objašnjena probabi-

listićka interpretacija unakrsne entropije i njezina povezanost s metodom maksimalne izglednosti (MLE). Na isti se način unakrsna entropija može interpretirati i u kontekstu neuronskih mreža: izlaz modela predstavlja procijenjene vjerojatnosti pripadnosti klasama, a minimiziranje unakrsne entropije odgovara maksimiziranju vjerodostojnosti opaženih oznaka. Kako bi se parametri mreže mogli prilagoditi tako da se vrijednost funkcije gubitka smanjuje, potrebno je izračunati gradijent gubitka u odnosu na svaki od njih. U dubokim mrežama gdje postoje milijuni (pa i stotine milijuna) parametara, nije efikasno gradijente za svaki parametar računati iz nule. Učinkovitiji pristup temelji se na pravilu ulančavanja 4.13, koje omogućuje rastavljanje složenih derivacija na umnožak jednostavnijih derivacija.

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial \mathcal{L}}{\partial x} \frac{\partial x}{\partial y}. \quad (4.13)$$

Budući da je neuronska mreža sastavljena od niza međusobno povezanih funkcija, gradijenti različitih parametara imaju zajedničke dijelove u svom izračunu. Zbog toga se gradijent funkcije gubitka može izračunati propagiranjem parcijalnih derivacija od izlaza mreže prema njezinim ulaznim slojevima, čime se izbjegava potreba za njihovim ponovnim računanjem. Ovaj postupak naziva se propagacija pogreške unatrag (engl. *backpropagation*).

Izračunati gradijenti se koriste za ažuriranje parametara mreže pri čemu način i jačina ažuriranja ovise o odabranom algoritmu optimizacije. U klasičnom gradijentom spustu parametri se ažuriraju u smjeru suprotnom od gradijenta funkcije gubitka, a veličina koraka određena je stopom učenja – hiperparametar koji određuje brzinu kretanja po prostoru parametara. Ne postoji univerzalno optimalna vrijednost stope učenja jer njezin prikladan izbor ovisi o samom zadatku. Premala stopa učenja može dovesti do vrlo sporog učenja, dok prevelika može uzrokovati osciliranje oko minimuma ili čak onemogućiti konvergenciju. U praksi se, osim klasičnog gradijentnog spusta, često koriste njegove varijante poput stohastičkog gradijentnog spusta (engl. *Stochastic Gradient Descent*, SGD) i adaptivnih metoda poput Adama (engl. *Adaptive Moment Estimation*). U ovom istraživanju koristi se AdamW koji uvodi odvojenu regularizaciju težina (radi sprečavanja prenaučenosti) od ažuriranja pomoću gradijenta što često omogućuje bolju kontrolu složenosti modela.

4.2.4. Objašnjivost

Gore objašnjeni algoritmi dubokog učenja mogu se prilagoditi velikim skupovima podataka i tako obavljati kompleksne zadatke, no oni spadaju u kategoriju *black-box* modela – nije jednostavno objasniti modelu odluku za dani ulazni primjer. Duboke neuronske mreže su zapravo zamršen skup težina podijeljene u slojeve. One kao takve ne pružaju intuitivnu logiku odlučivanja. U suprotnu kategoriju spadaju tzv. *white-box* modeli poput linearne regresije i stabala odluke, gdje se iz same konstrukcije algoritma (koeficijenti uz značajke kod linearne regresije, pravila grananja kod stabala) vidi objašnjenje.

Veća objašnjivost modela nerijetko je popraćena lošijim rezultatima u usporedbi s *black-box* modelima, pogotovo kad se radi o kompleksnijim podacima. Unatoč tome, interes za povećanje objašnjivosti modela kontinuirano raste, a za to postoji više razloga. Jedan od njih je povećanje povjerenja korisnika u model, zbog čega su korisnici spremniji prihvatiti proizvedenu tehnologiju. Osim toga, razvojem umjetne inteligencije nameće se važan etički problem pristranosti modela. Algoritmi umjetne inteligencije su jako osjetljivi na skup podataka koji imaju na raspolaganju. Ako je zbog nekog razloga, kao što je slučaj kod npr. starijih skupova podataka, on pristran, model može usvojiti postojeće obrasce pristranosti i prenijeti ih na svoje predikcije. Time model postaje pristran, iako sama pristranost nije izravno zadana u algoritmu. Primjer takvog problema jest slučaj Facebookova algoritma koji je crne osobe klasificirao kao primate, što je povezano s nedovoljnom reprezentacijom crnih osoba u podacima korištenima za učenje [20]. Takve pogreške mogu dovesti do toga da algoritam nesvjesno diskriminira određenu skupinu ljudi, dok bez mogućnosti uvida u način donošenja odluke to nije moguće jednostavno otkriti.

Posebno je važna objašnjivost u područjima poput medicine jer se radi o ljudskom zdravlju. U takvom je kontekstu važno razumjeti zašto model, primjerice, predviđa prisutnost tumora ili preporučuje određenu terapiju. S objašnjivošću je povezan i problem odgovornosti: u slučaju pogrešne odluke potrebno je moći utvrditi kako je do nje došlo i tko snosi odgovornost. Zbog toga se stavlja naglasak na razvijanje metoda koje objašnjavaju *black-box* modele. Jedna od takvih metoda, namijenjena upravo konvolucijskim neuronskim mrežama, jest Grad-CAM (engl. *Gradient-weighted Class Activation Mapping*).

Svrha metode objašnjenja modela Grad-CAM je odrediti koji su pikseli ulazne slike najviše doprinijeli odluci modela. Metoda koristi gradijente izlaza koji odgovara odabranoj klasi u odnosu na aktivacije posljednjeg konvolucijskog sloja. Za odabranu klasu c , najprije se računaju gradijenti izlaza modela y^c u odnosu na aktivacije A^k k -te aktivacijske mape posljednjeg konvolucijskog sloja:

$$\frac{\partial y^c}{\partial A_{ij}^k}. \quad (4.14)$$

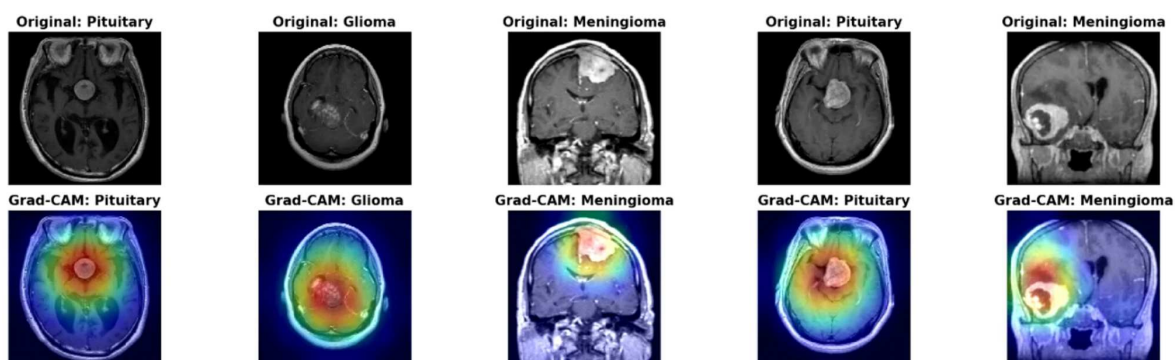
Ti se gradijenti zatim usrednjavaju preko prostornih dimenzija aktivacijske mape kako bi se dobila težina α_k^c , koja predstavlja važnost k -te aktivacijske mape za odabranu klasu:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \quad (4.15)$$

gdje je Z broj prostornih elemenata aktivacijske mape. Grad-CAM mapa za klasu c tada se dobiva kao ponderirana suma aktivacijskih mapa. Vrijednosti u toj sumi mogu biti i pozitivne i negativne. Pozitivne vrijednosti označavaju područja koja podržavaju predikciju klase c , dok negativne predstavljaju područja koja je potiskuju. S obzirom na to da Grad-CAM metoda objašnjenja modela želi vizualizirati upravo ono što je model koristio kao dokaz za svoju odluku, propuštaju se samo pozitivne vrijednosti pomoću ReLU:

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right). \quad (4.16)$$

Budući da je mapa prostorno manje rezolucije od ulazne slike, interpolacijom se povećava na njezine dimenzije te se može prikazati kao toplinska mapa preko izvorne slike. Na resultantnom prikazu crvena i žuta područja označavaju dijelove slike koji su najviše pridonijeli predikciji odabrane klase, dok plava područja predstavljaju manji ili zanemariv doprinos. Na slici 4.7. to se dobro vidi – gornji red prikazuje izvorne MRI snimke, dok donji red prikazuje odgovarajuće Grad-CAM toplinske mape.



Slika 4.7. Primjer Grad-CAM aktivacija. Preuzeto s [21]

5. Primjena umjetne inteligencije u neuropsihijatriji

Znanstveni radovi na području psihologije i strojnog učenja se sve češće usredotočuju na automatiziranu dijagnostiku i istovremeno otkrivaju prepreke u prepoznavanju mentalnih poremećaja. Preciznije što se tiče bipolarnog poremećaja i shizofrenije, brojni pristupi koriste podatke iz EEG-a i MRI-a kako bi se razvili klasifikacijski modeli sposobni razlikovati navedene poremećaje od drugih psihijatrijskih stanja, poput unipolarne depresije. Međutim, istraživanja pokazuju da se različiti poremećaji često anatomske i funkcionalno preklapaju, što dodatno otežava točnu dijagnozu. U nastavku su analizirani odabrani znanstveni radovi koji se bave primjenom strojnog i dubokog učenja u otkrivanju bipolarnog poremećaja i/ili shizofrenije, pri čemu se posebno razmatraju korištene metode, točnost klasifikacije te praktične prepreke i ograničenja.

Jedno od takvih istraživanja provedeno je 2021. godine. Autori su koristili dotad najveći skup podataka [22] koji se sastoji od MR slika 3020 pacijenata, od kojih 853 pati od bipolarnog poremećaja. Podaci su prikupljeni s više središta (njih 13). Ovakav pristup dovodi do različitosti u podacima zbog raznih faktora, poput različitih postavki uređaja korištenih za snimanje i različitih tipova poremećaja (5 od 13 centara je imalo samo bipolarni poremećaj tip 1, 1 centar samo tip 2, ostali su imali spektar). Uz to, sam bipolarni poremećaj ima varijabilne simptome što dodatno doprinosi heterogenosti podataka. To nije nužno loše zato što to omogućava algoritmu strojnog učenja da pronađe općenite uzorke povezane s poremećajem pa model može bolje generalizirati. Nisu korištene sirove T1-mjerene MR slike, već alat FreeSurfer kojim se, po uzoru na atlas Desikan–Killiany, svaka hemisfera dijeli na 7 subkortikalnih i 34 kortikalne regije (engl. *region of interest*, ROI). Za svaku od kortikalnih regija izdvojene su njihova kortikalna debljina i površina, a za subkortikalne njihov volumen – ukupno 150 značajki. Također je uzeta i značajka in-

trakranijalni volumen. Uvijek se koristio stroj potpornih vektora s linearnom jezgrom, a isprobano je više klasifikatora ovisno o podjeli podataka. Najbolji rezultati su postignuti kad su korišteni podaci iz svih centara za učenje uz unakrsnu validaciju. S takvim je modelom postignuta točnost od 65,23% i ROC-AUC mjera od 71,49%. Rezultati nisu dovoljni za kliničku upotrebu (bar 80%). Moguće poboljšanje bilo bi koristiti duboko učenje koje bi imalo pristup sirovim slikama, budući da kod obrade FreeSurferom dolazi do neizbježnog gubitka informacija. SVM s linearnom jezgrom omogućuje uvid u važnosti značajki, a u ovom istraživanju važne su značajke bile u skladu s teorijskom pozadinom – regije poput amigdale i hipokampusa su se našle u skupu najvažnijih značajki. Klasifikacija nije ovisila o jednom centru – kad su iz modela izostavili podatke s lokacija koje su imale najbolje rezultate (u varijanti učenja 13 klasifikatora, svaki s podacima iz vlastitog centra), ukupna točnost klasifikacije nije znatno pala. Značajke koje su pridonijele klasifikaciji u agregiranom uzorku bile su vrlo slične onima iz najboljeg pojedinačnog centra (Cohenova kappa = 0,83). Konačno, skup podataka bio je jako heterogen. Također je primijećeno da su starija dob i terapija antiepilepticima bili povezani s većom vjerojatnošću točne klasifikacije, dok spol nije imao neki utjecaj. Zaključno, ovaj rad predstavljao je korak naprijed u primjeni strojnog učenja za detekciju bipolarnog poremećaja. Iako razina točnosti nije dovoljna za kliničku primjenu, isticala se mogućnost poboljšanja primjenom naprednijih metoda (npr. dubokog učenja) na sirove podatke.

Drugi rad [23] bavio se detekcijom psihoze, s fokusom na razlikovanje kontrolne grupe, pacijenata sa shizofrenijom i pacijenata s bipolarnim poremećajem. Također su se koristili podaci dobiveni iz MR slika, ali osim ROI koristile su se i npr. metode VBM (engl. *voxel-based morphometry*) i WBM (engl. *Wavelet-based morphometry*). Voksel predstavlja najmanju prostornu jedinicu unutar MR slike, dok metode poput VBM-a omogućuju analizu volumena sive tvari na razini pojedinačnih vokselâ, bez prethodnog grupiranja u regije interesa. Na taj se način može zadržati veća količina prostorne informacije sadržane u MR podacima. U radu [23] ispitano je osam različitih klasifikacijskih algoritama: ridge, lasso, elastic net i L0 norma logističke regresije, stroj potpornih vektora, regularizirana diskriminantna analiza, klasifikator s Gaussovim procesom i slučajna šuma. Zanimljivo je da odabir algoritma nije značajno utjecao na točnost klasifikacije, dok se odabir vrste značajki pokazao ključnim – upravo su modeli koji su koristili VBM značajku imali najvišu klasifikacijsku točnost: 63% u razlikovanju zdravih i bipolarnih osoba, 75%

u razlikovanju zdravih i shizofrenih osoba te 62% u razlikovanju bipolarnih i shizofrenih osoba. Autori su pokušali povećati točnost kombiniranjem svih značajki u jedan model, ali bilo je potrebno reducirati njihov ukupan broj pomoću PCA. Takav model je čak imao lošiju točnost. Rad posebno upućuje na probleme već spomenute pogrešne dijagnoze; ovdje konkretno model najlošije diferencira dvije bolesne osobe. U istraživanju su ispitane i višeklasne klasifikacije koristeći pristupe one-vs-one i one-vs-all. Iako su ovi pristupi omogućili istovremeno razlikovanje sve tri skupine (kontrole, bipolarni poremećaj, shizofrenija), odziv klasifikacije za bipolarni poremećaj bio je nizak i približno na razini slučajnog pogađanja (oko 33%).

Navedeni rezultati bili su posebno značajni za oblikovanje metodologije ovog istraživanja. Budući da se koristi isti skup podataka, u istraživanju se slijede slični principi – ispituju se iste vrste značajki, primjenjuju se slični klasifikacijski algoritmi te se njihovi rezultati međusobno uspoređuju. Na taj se način rezultati dobiveni u ovom istraživanju mogu promatrati u kontekstu prethodnog rada, pri čemu je moguće uspoređivati rezultate dobivene primjenom sličnih metodoloških pristupa na istom skupu podataka.

Još jedan pristup predstavljen je u radu [24], u kojem su autori pokušali istovremeno razlikovati zdrave osobe, pacijente sa shizofrenijom, bipolarnim poremećajem i shizoafektivnim poremećajem. Za razliku od prethodno opisanih istraživanja, korišteni su podaci dobiveni funkcionalnom magnetskom rezonancijom mozga u stanju mirovanja (engl. *resting-state functional magnetic resonance imaging*, rs-fMRI). U istraživanju je korišteno 693 ispitanika iz različitih centara, od kojih je 229 bilo zdravih, 176 sa shizofrenijom, 140 s bipolarnim poremećajem i 129 sa shizoafektivnim poremećajem. Podaci korišteni u istraživanju su dobiveni temeljitom predobradom sirovih slika. Iz njih je prvo kombinacijom PCA i ICA metoda izdvojeno 100 nezavisnih komponenti, a zatim je pomoću GIG-ICA, koristeći grupne komponente kao vodič, za svakog ispitanika rekonstruirana individualna verzija tih komponenti. Time su za svakog ispitanika izdvojena 54 vremenska niza (engl. *time courses*, TC), nakon uklanjanja komponenti povezanih s fiziološkim šumom i pokretima. Kao glavni klasifikator autori su predložili višeskalnu rekurentnu neuronsku mrežu (engl. *multi-scale recurrent neural network*, MsRNN), koja može iskoristiti vremensku strukturu podataka i učiti složenije obrasce iz fMRI vremenskih nizova. Najprije su korištene 1D konvolucije s filtrima različitih vremenskih širina

(2, 4 i 8), čime su izdvojene značajke na različitim vremenskim skalama. Na kraju su korištene dvije GRU (engl. *gated recurrent unit*) ćelije da se riješi problem nestajućeg gradijenta. Za usporedbu su korišteni i klasični algoritmi strojnog učenja, odnosno SVM, AdaBoost i slučajna šuma. U četveroklasnoj klasifikaciji MsRNN je ostvario točnost od 46%, što je bilo značajno bolje od slučajnog pogađanja i bolje od korištenih klasičnih metoda. Ipak, rezultat pokazuje koliko je težak problem istovremenog razlikovanja više psihijatrijskih poremećaja, osobito zbog velikog preklapanja njihovih simptoma. Kad se klasifikacija ograničila na bipolarni poremećaj i zdravu kontrolnu skupinu, ostvarena je znatno viša točnost od 71,6%. Autori su dodatno pokušali interpretirati naučeni model primjenom pristupa *leave-one-feature-out*, pri čemu su pojedinačne funkcionalne komponente izostavljane iz modela kako bi se procijenila njihova važnost pri klasifikaciji. Kod četveroklasne klasifikacije među najvažnijim područjima istaknuli su se hipokampus, suplementarno motorno područje i paracentralni režanj, dok se primjerice mali mozak posebno izdvojio pri razlikovanju bipolarnog poremećaja od zdrave kontrolne skupine. Ovaj je rad posebno relevantan za prethodno opisane pristupe jer pokazuje potencijal dubokog učenja za izravno iskorištavanje vremenskih informacija sadržanih u fMRI podacima, umjesto oslanjanja isključivo na unaprijed definirane značajke. Istovremeno, relativno niska točnost četveroklasne klasifikacije pokazuje da veća složenost modela sama po sebi nije dovoljna za pouzdano razlikovanje psihijatrijskih poremećaja, osobito kada među njima postoji značajno preklapanje u obrascima moždane aktivnosti.

Zadnji rad [25] koji treba spomenuti objavljen je 2020. godine, a bavi se detekcijom shizofrenije pomoću dubokog učenja. Autori su koristili poveći skup podataka – sastavljen od pet zasebnih skupova iz pet različitih centara, rezultirajući skupom od 866 strukturnih MR slika, od kojih je 443 pripadalo pacijentima sa shizofrenijom, a 423 zdravim osobama. Kao zaseban testni skup korišteno je 60 slika iz jednog centra. Taj se skup razlikovao od podataka korištenih za učenje, između ostalog po dobi i omjeru spolova, što su autori i ciljali jer su tako mogli ispitati može li model generalizirati na populaciju s drugačijim karakteristikama; je li model zapravo naučio korisne obrasce ili je naučio neke specifičnosti uređaja/pacijenata. Za ulaz u model nisu izravno koristili cijele 3D NIfTI slike, već su pojedinačne transverzalne presjeke spojili u videozapis. Taj je pristup potaknut ljudskom obradom MR slika – sekvencijalni pregled pojedinačnih presjeka. Za klasifikaciju je korištena 3D CNN arhitektura s četiri konvolucijska sloja, pri čemu su

korištene jezgre dimenzija $3 \times 3 \times 3$, a nakon svakog konvolucijskog sloja uslijedio je sloj sažimanja maksimumom. Na kraju dolazi gusto povezani sloj. Kako bi se smanjila mogućnost prenaučivosti, autori su primijenili deseterostruku unakrsnu validaciju, a učenje su zaustavili nakon postizanja konvergencije. Dobiveni rezultati bili su uspoređeni s rezultatima jednostavne logističke regresije koja je koristila prosječan intenzitet slike kao prediktor. Rezultati su pokazali da korištenjem dubokog učenja uspijevaju dobiti točnost na testnom setu od 70%, dok logistička regresija uspijeva dobiti samo 51%, dakle intenzitet i kvaliteta slike nisu dovoljni za klasifikaciju. Autori su također istražili koje regije mozga najviše znače za predikciju – najveći pad AUC-a, s 0,96 na 0,57, zabilježen je nakon uklanjanja regije koja približno odgovara desnom temporalnom području, što se slaže s pronalascima iz ranijih istraživanja. Autori ističu da su u testnom skupu pacijenti bili mlađi, što u startu već otežava dijagnozu jer postoji mogućnost da se strukturne promjene povezane sa shizofrenijom još nisu razvile.

6. Skup podataka

Vrlo važan i obavezan korak u bilo kakvom algoritmu temeljenom na strojnom učenju jest prikupljanje podataka. Podataka danas ima u izobilju, ali je izazov pronaći dovoljno kvalitetnih podataka, specifičnih za dani problem. Kvaliteta podataka je ta koja može značajno pomoći, odnosno odmoći, u postizanju dobrih rezultata. Zbog čega je tako? Budući da algoritam mora učiti na podacima koji su prikupljeni, oni moraju točno reprezentirati područje istraživanja. Algoritam sam za sebe ne zna ništa i sve što će naučiti bit će temeljeno na podacima koji su prikupljeni, stoga ako se koriste loši podaci, algoritam će se pokušati njima prilagoditi te će izvući neželjene zaključke. Primjeri loših karakteristika skupova podataka su krivo označeni podaci, neuravnoteženost (dominacija primjera određene klase) ili skupovi s nedovoljnim brojem podataka. Taj zadnji problem često zahvaća istraživanje poput ovoga – biomedicinsko područje. Mnogi zakoni, poput GDPR-a, štite pacijentovu privatnost i otežavaju prikupljanje zdravstvenih podataka. Zbog navedenog se često provodi augmentacija podataka – primjenom manjih, očekivanih varijacija umjetno se povećava broj podataka u skupu. Taj je pristup isproban i u ovom istraživanju.

Skup podataka za ovo istraživanje [26] prikupljen je u Benito Menni CASM i Mare de Déu de la Mercè bolnicama u Španjolskoj. Sastoji se od 373 subjekta, od kojih 128 pripada zdravoj kontrolnoj grupi, 128 grupi s bipolarnim poremećajem tip 1, a 127 pacijenata pati od shizofrenije. Kontrolnu grupu čini nemedicinsko osoblje spomenutih bolnica, ne uzimajući u obzir osoblje koje ima povijest mentalnih bolesti ili liječenja psihotropičkim lijekovima. Svi su podaci dobiveni koristeći isti uređaj za magnetsku rezonanciju mozga jačine 1.5-T, a prikupljene su T1-mjerene strukturne MR slike. Budući da skup podataka ne sadrži sirove MR slike, za potrebe klasifikacije iz njih su prethodno izdvojene i obrađene sljedeće skupine značajki:

1. Kortikalna debljina (CT) – dobivena je obradom strukturnih MR slika pomoću alata FreeSurfer. U procesu obrade najprije su uklonjena tkiva koja nisu dio mozga, a zatim su određene granice između sive i bijele tvari te površina moždanog korteksa. Na temelju tih granica izračunata je debljina korteksa u pojedinim vrhovima kortikalne površine. Izdvojene su zasebne mape za desnu i lijevu hemisferu.
2. Kortikalni volumen (CV) – Uz CT, pomoću alata FreeSurfer izračunata je i površina korteksa, koja uz CT služi za izračun mape kortikalnog volumena za obje hemisfere. Sve 4 mape (CT i CV) su dodatno zaglađene koristeći Gaussov filter.
3. Volumetrijske morfometrijske značajke sive i bijele tvari (VBM) – strukturne slike najprije su segmentirane na sivu i bijelu tvar, a zatim prostorno normalizirane prema MNI152 standardu kako bi se omogućila usporedba anatomskih struktura različitih ispitanika. Rezultantne mape su smanjene na rezoluciju 4x4x4 mm radi smanjenja računalnih troškova.
4. Valične morfometrijske značajke sive i bijele tvari (WBM) – izdvajaju se primjenom valične transformacije na VBM podatke. Na taj se način izvorni VBM podaci transformiraju u skup valičnih koeficijenata, a zadržani su oni koeficijenti koji objašnjavaju najveći postotak energije.
5. Volumeni regija interesa (ROI) – Korištenjem alata FreeSurfer mozak je podijeljen na regije interesa, a izvučeni su prosječni volumeni tih regija, kao i malog mozga, bijele tvari i moždanih ventrikula. Također, nakon standardizacije navedenih podataka, kao dodatne značajke izdvojene su interakcije između svih regija.

7. Implementacija

7.1. Klasično strojno učenje

U prethodnom poglavlju spomenute su značajke korištene za učenje modela strojnog učenja, uključujući i WBM značajke. Budući da su one dobivene vlastitom implementacijom, u nastavku je detaljnije opisan postupak njihove ekstrakcije iz VBM podataka.

Postupak ekstrakcije značajki temeljen je na pristupu opisanom u radu [27]. U ovom radu implementiran je prilagođeni postupak 3D diskretne ortogonalne transformacije valićima (3D-DOWT), pri čemu su korišteni ortogonalni i simetrični Battle–Lemarié splajn valići, kao u izvornom WBM pristupu. Postupak se sastoji od više koraka; ukratko započinje zaglađivanjem VBM slika Gaussovom jezgrom širine 7mm. Iz tih tako zaglađenih slika izračunata je globalna prosječna VBM slika. Ona je, po uzoru na rad [27], korištena za određivanje skupa najinformativnijih koeficijenata. Na njoj je provedena višerazinska 3D-DOWT. U ovoj implementaciji transformacija valićima izvedena je izravnim filtriranjem i poduzorkovanjem u prostornoj domeni. Jednodimenzionalna transformacija primijenjena je zasebno duž svake od triju prostornih dimenzija. U svakoj dimenziji signal se filtrira niskopropusnim (L) i visokopropusnim (H) Battle–Lemarié filtrom, nakon čega se rezultat poduzorkuje za faktor dva. Kombinacijom L i H filtara, dobiva se osam podprostora za jednu razinu, a kao ulaz za sljedeću razinu transformacije uzima se kombinacija LLL. Koeficijenti svih osam podprostora na svakoj razini konkatenuju se u konačni vektor valićnih koeficijenata. Isječak koda kojim je implementiran opisan postupak prikazan je na slici 7.1. Iz njega se metodom minimalne duljine opisa (*Minimum Description Length*, MDL) zadržavaju najinformativniji koeficijenti prema kriteriju opisane energije [28] – za svaki mogući broj k zadržanih koeficijenata izračunava se preostala energija signala nakon uklanjanja energije k najvećih koeficijenata. Na temelju broja zadržanih koeficijenata i preostale energije računa se vrijednost MDL kriterija. Po-

zicije najinformativnijih koeficijenata su zapamćene te se od svakog ispitanika zasebno uzimaju njegovi koeficijenti na odgovarajućim mjestima. Isječak koda vidljiv je na slici 7.2. Tako su za svakog ispitanika dobivene WBM značajke. Radi smanjenja vremenske složenosti pri učenju modela, koeficijenti za svakog ispitanika su unaprijed izračunati i pohranjeni, a tijekom izvođenja se računaju samo koeficijenti za prosječnu sliku.

```
def dwt_3d(levels, mean_image, h, g): 1 usage
    new_approx = mean_image
    coeffs_list = []
    for i in range(levels):
        coeffs = dwt3d(new_approx, h, g)
        flat_vector = np.concatenate([v.ravel() for v in coeffs.values()])
        coeffs_list.append(flat_vector)
        new_approx = coeffs['LLL']
    return np.concatenate(coeffs_list)
```

Slika 7.1. Isječak koda za višerazinsku 3D diskretnu transformaciju valićima i izdvajanje dobivenih koeficijenata.

```
idx_sorted = np.argsort(np.abs(d))[:,::-1]
d_sorted = d[idx_sorted]

total_energy = np.sum(d**2)
residual_energy = np.empty(K + 1)
residual_energy[0] = total_energy
for k in range(1, K + 1):
    residual_energy[k] = residual_energy[k - 1] - d_sorted[k - 1]**2

eps = np.finfo(float).eps # avoid log(0)
residual_energy = np.maximum(residual_energy, eps)

k_vals = np.arange(K + 1)
amd1 = (
    1.5 * k_vals * np.log2(N)
    + 0.5 * N * np.log2(residual_energy)
)

k_hat = np.argmin(amd1)
```

Slika 7.2. Isječak koda za odabir valićnih koeficijenata primjenom kriterija minimalne duljine opisa.

Budući da je cilj bio ispitati koje kombinacije značajki daju najbolje rezultate klasifikacije, proces učenja uključivao je unakrsnu validaciju (engl. *cross-validation*, CV) i postupak pohlepnog pretraživanja unaprijed (engl. *greedy forward search*, GFS).

Proces unakrsne validacije, osim klasične podjele na skup za učenje, validaciju i test, uključivao je više takvih podjela. Naime, slučajna podjela malog skupa podataka može rezultirati nereprezentativnim skupovima, što može dovesti do precjenjivanja ili podcjenjivanja stvarnih performansi modela. Zbog toga se više puta, u ovom radu konkretno pet puta, ponavlja proces podjele na skupove kako bi se smanjio utjecaj same podjele i dobila vjerodostojnija procjena usrednjavanjem rezultata po pojedinim ponavljanjima (engl. *fold*).

Budući da bi ispitivanje svakog postojećeg podskupa značajki bilo računalno neizvedivo (2044 kombinacije, uz još i ugniježdenu unakrsnu validaciju), prostor pretraživanja smanjen je pomoću postupka pohlepnog pretraživanja unaprijed. Postupak započinje uzimanjem najboljeg pojedinačnog skupa značajki te mu se u svakoj sljedećoj iteraciji dodaje skup koji u kombinaciji s prethodno odabranima daje najbolje performanse modela. Za procjenu performansi korištena su dva klasifikacijska modela: logistička regresija s L2 regularizacijom i vrijednošću parametra $C=1$ te slučajna šuma, a kao rezultat pojedinog modela se uzima prosječna točnost testnih podskupova. Za sljedeću iteraciju odabire se značajka koja ostvaruje najbolji rezultat prema boljem rezultatu od ta dva modela. Odgovarajući isječak koda vidljiv je na slici 7.3. U praksi se pokazalo da performanse ili degradiraju dodavanjem značajki ili ostaju približno jednake, stoga se postupak završava nakon odabira pet značajki.

U sklopu optimizacije korišten je pristup ugniježdene unakrsne validacije (engl. *Nested Cross-Validation*) – za svaku usporedbu provedena je vanjska podjela na pet preklopa (engl. *fold*) kojom se nastojalo dobiti nepristranu procjenu generalizacije modela te unutarnja podjela na tri preklopa, koja je služila za odabir najbolje kombinacije hiperparametara. Unutar navedenog postupka ispitana su tri klasifikacijska modela: logistička regresija s L2 regularizacijom, kvadratna diskriminantna analiza i slučajna šuma. Za logističku regresiju optimiran je parametar regularizacije C , s vrijednostima 0,01, 0,1 i 1, dok je kod QDA modela optimiran parametar regularizacije s vrijednostima 0, 0,2 i 0,5. Za slučajnu šumu nisu dodatno optimirani hiperparametri modela, već su korištene unaprijed definiranim vrijednosti za parametre zbog činjenice da slučajne šume često daju dobre rezultate bez optimiranja hiperparametara [29]. Za sva tri modela dodatno je ispitivan postupak redukcije dimenzionalnosti, pri čemu su uspoređene PCA redukcija

```

remaining = list(feature_names)
selected = []

while remaining:
    best_candidate = None
    best_result = None

    for fname in remaining:
        combo = tuple(selected + [fname])
        result = evaluate_combo(combo, X_pair, y_pair, SCREEN_MODELS)
        best_score = max(result.values())

        if best_result is None or best_score > max(best_result.values()):
            best_candidate = fname
            best_result = result

    selected.append(best_candidate)
    remaining.remove(best_candidate)
    print(f"Selected so far: {selected} -> {best_result}")

```

Slika 7.3. Isječak koda korišten za pohlepan odabir skupova značajki.

i redukcija na temelju F-vrijednosti (engl. *F-value*). Broj zadržanih značajki ovisio je o korištenoj metodi: za logističku regresiju i slučajnu šumu ispitane su vrijednosti 10, 100 i 500 značajki, dok su za QDA korištene vrijednosti 10, 50 i 100 značajki. Isječak koda za učenje prikazan je na slici 7.4. Odabrani skupovi značajki se objedinjuju u jedan skup pomoću *FeatureUnion*. Dobivene značajke zatim se prosljeđuju klasifikacijskom modelu unutar strukture cjevovoda. Optimalna kombinacija hiperparametara određuje se pretraživanjem po rešetci (*GridSearchCV*), uz unutarnju unakrsnu validaciju. Kao primjer cjevovoda, na slikama 7.5. i 7.6. prikazana je implementacija cjevovoda za učitavanje i predobradu VBM podataka. *VBM_load* učitava VBM značajke za ispitanike uključene u trenutačni skup podataka. Na skupu za učenje određuje se maska koja se zatim primjenjuje pri učitavanju podataka, nakon čega se dobivene značajke standardiziraju, što je definirano redoslijedom koraka u funkciji *make_vbm_wm_pipeline*. Za svaku vrstu značajki definiran je zaseban cjevovod koji objedinjuje korake potrebne za pripremu podataka prije njihove primjene u klasifikacijskom modelu.

7.2. Duboko učenje i objašnjivost

Za klasifikaciju pomoću dubokog učenja isprobana su dva pristupa – učenje mreža od početka (engl. *from scratch*) i učenje već prednaučenih mreža. Pristupi imaju sličnosti u

```

# creating new pipeline (calling the method passed as dictionary value)
# feature_dict is a dictionary of feature:function_for_feature_pipeline pairs
features = FeatureUnion([
    (feature_name, feature_dict[feature_name]()) for feature_name in combo
])

# LinearSVC is written because python needs to know model has a .score method
pipe = Pipeline([
    ("features", features),
    ("dimred", "passthrough"),
    ("model", LinearSVC())
], memory=memory)

param_grid = make_param_grid()
search = GridSearchCV(
    pipe,
    param_grid=param_grid,
    cv=inner_cv,
    n_jobs=1,
    verbose=2
)

search.fit(X_train_idx, y_train)
fold_parameters.append(search.best_params_)

```

Slika 7.4. Isječak koda za učenje.

```

def make_vbm_wm_pipeline():
    return Pipeline([
        ("load", VBM_load("/content/drive/MyDrive/zenodo_dataset/vbm_wm")),
        ("scaler", StandardScaler())
    ])

```

Slika 7.5. Isječak koda koji prikazuje cjevovod, odnosno pripremu VBM podataka prije njihovog korištenja.

učenju:

- Ulaz u mrežu predstavljaju 3D VBM strukturne slike. Posebno su učene mreže sa slikama sive tvari i posebno sa slikama bijele tvari.
- Učenje mreža provedeno je korištenjem optimizatora Adam s regularizacijom težina.
- Model je učen do najviše sto epoha. S ciljem smanjenja prenaučenosti, nakon svake se epohe izračunavao gubitak na skupu za validaciju. Ako ne bi došlo do smanjenja gubitka na skupu za validaciju kroz petnaest epoha, učenje je završeno (engl. *early stopping*), a uzet je model s najmanjim gubitkom na validacijskom skupu. Radi provjere, prikazan je graf kretanja validacijskog gubitka kroz epohe. To omogu-

```

class VBM_Load(BaseEstimator, TransformerMixin):
    def __init__(self, folder_name):
        self.folder_name = folder_name
        self.mask = None

    def fit(self, X, y=None):
        filenames = [all_subjects[ind] + ".nii.gz" for ind in X]
        self.mask = compute_mask(self.folder_name, filenames)
        return self

    def transform(self, X):
        filenames = [all_subjects[ind] + ".nii.gz" for ind in X]
        return load_img_data_with_mask(self.folder_name, filenames, self.mask)

```

Slika 7.6. Isječak koda za učitavanje VBM značajki.

čuje detaljniji uvid u ponašanje modela – na temelju grafa može se procijeniti je li smanjenje gubitka stabilno i praćeno daljnjim poboljšanjem ili se javlja izolirano, nakon čega se validacijski gubitak ponovno povećava.

- Kao i kod klasičnog strojnog učenja, korišten je pristup peterostruke unakrsne validacije.

7.2.1. Učenje od početka

U prvoj skupini eksperimenata uspoređene su četiri arhitekture koje se razlikuju prema broju konvolucijskih i potpuno povezanih slojeva: model s jednim konvolucijskim i jednim potpuno povezanim slojem, model s jednim konvolucijskim i dva potpuno povezana sloja, model s dva konvolucijska i jednim potpuno povezanim slojem te model s dva konvolucijska i dva potpuno povezana sloja. U svim navedenim arhitekturama nakon konvolucije primijenjeni su ReLU aktivacija i sažimanje maksimumom, a nakon posljednjeg konvolucijskog sloja i sloj globalnog prosječnog sažimanja, čime je prostorni prikaz sveden na jednu vrijednost po kanalu prije klasifikacije.

U drugoj skupini eksperimenata ispitivan je utjecaj vrste normalizacije. U odnosu na osnovni model bez normalizacije, ispitane su varijante s normalizacijom mini-skupa (engl. *batch normalization*) i normalizacijom po grupama. Dok normalizacija mini-skupa, kao što joj i ime kaže, računa srednju vrijednost i varijancu primjera unutar iste mini grupe, normalizacija po grupama računa te vrijednosti zasebno za svaki primjer, pri čemu se kanali dijele u grupe.

U trećoj skupini eksperimenata ispitivan je utjecaj položaja normalizacije u arhitekturi. Konkretno, ispitane su dvije mogućnosti položaja normalizacije mini-skupa: između konvolucijskog sloja i aktivacije te nakon aktivacije. Ova je usporedba provedena jer, iako je uobičajen položaj normalizacije prije sloja aktivacije, ističe se mogućnost povećanja točnosti nakon njezinog pomicanja poslije aktivacije [30]. U tom radu navedeno je kako je razlika uzrokovana položajem normalizacije bila nezamjetna kad se koristila ReLU aktivacija, dok je razlika bila značajna kad se koristila druga aktivacija, poput tangensa hiperbolnog. Zato je u ovom eksperimentu umjesto ReLU korišten upravo tangens hiperbolni.

U četvrtoj skupini eksperimenata ispitivan je utjecaj prostornog sažimanja značajki. U tu su svrhu uspoređene dvije inačice iste osnovne arhitekture – jedan konvolucijski sloj, nakon kojeg slijedi ReLU i sažimanje maksimumom. U prvoj se nakon sažimanja koristi globalno prosječno sažimanje, dok se u drugoj prostorne značajke izravnavaju u jedan vektor. Time se ispituje utjecaj zadržavanja prostorne informacije na klasifikacijsku uspješnost. Dok globalno prosječno sažimanje znatno smanjuje broj značajki i parametara klasifikacijskog sloja, drugi pristup zadržava prostornu strukturu značajki, ali rezultira većim brojem parametara i mogućim *overfittingom*.

Za svaku od navedenih arhitektura isprobano je više vrijednosti za iduće hiperparametre:

- Broj jezgri u konvolucijskim slojevima
- Veličina jezgre u konvolucijskim slojevima
- Broj neurona u potpuno povezanim slojevima
- Veličina regije sažimanja u slojevima sažimanja
- Veličina grupe u slojevima normalizacije po grupi

Isprobane konfiguracije vidljive su na slikama 7.7. i 7.8. Isprobana je svaka kombinacija hiperparametara.

```

# Experiment group 1
experiment_group_1_configs = {
    "model_1_conv_1_fc": [BrainCNN3D_1conv_1FC, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
    "model_1_conv_2_fc": [BrainCNN3D_1conv_2FC, {
        "n_channels": [[8, 4], [16, 8], [32, 16]],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
    "model_2_conv_2_fc": [BrainCNN3D_2conv_2FC, {
        "n_channels": [[4, 8, 4], [8, 16, 8], [16, 32, 16]],
        "kernel_size": [[3, 3], [5, 5], [5, 3]],
        "maxpool_size": [[2, 2], [4, 4], [2, 4]],
    }],
    "model_2_conv_1_fc": [BrainCNN3D_2conv_1FC, {
        "n_channels": [[4, 8], [8, 16], [16, 32]],
        "kernel_size": [[3, 3], [5, 5], [5, 3]],
        "maxpool_size": [[2, 2], [4, 4], [2, 4]],
    }],
}

# Experiment group 2
experiment_group_2_configs = {
    "model_batch": [BrainCNN3D_1conv_batch, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
    "model_group": [BrainCNN3D_1conv_group, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
        "num_groups": [4, 8, 16]
    }],
}

```

Slika 7.7. Konfiguracije za prvu i drugu skupinu

7.2.2. Prednaučene mreže

Kako sami skup podataka ima malo primjera, osim učenja mreže od početka, isproban je i pristup prijenosnog učenja, odnosno finog podešavanja. Kod prijenosnog učenja koriste se unaprijed naučene duboke neuronske mreže, a zatim se njihovi naučeni parametri koriste za ekstrakciju značajki u ranijim slojevima, a novododani zadnji dio se prilagođava specifičnom skupu podataka. Fino podešavanje temelji se na sličnom principu, samo što se u tom slučaju dalje nastavlja učiti cijela mreža ili dio nje. Cilj je ispitati može li znanje naučeno na većim medicinskim skupovima podataka poboljšati klasifikaciju na promatranom skupu.

```

# Experiment group 3
experiment_group_3_configs = {
    "model_before": [BrainCNN3D_1conv_batch_before, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
    "model_after": [BrainCNN3D_1conv_batch_after, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
}

# Experiment group 4
experiment_group_4_configs = {
    "model_GAP": [BrainCNN3D_1conv_1FC_GAP, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
    }],
    "model_Flatten": [BrainCNN3D_1conv_1FC_Flatten, {
        "n_channels": [8, 16, 32],
        "kernel_size": [3, 5],
        "maxpool_size": [2, 4],
        "input_shape": [(46, 55, 46)]
    }],
}

```

Slika 7.8. Konfiguracije za treću i četvrtu skupinu

Kao osnovne arhitekture korištene su dvije 3D ResNet mreže iz projekta MedicalNet [31], odnosno ResNet-10 i ResNet-18. Obje arhitekture pripadaju obitelji rezidualnih neuronskih mreža (engl. *Residual Networks*), koje koriste rezidualne veze (tzv. skip veze) kako bi se olakšalo učenje dubljih mreža. Razlika između korištenih modela prvenstveno proizlazi iz njihove dubine i broja parametara. ResNet-18 predstavlja složeniji model te ima veći kapacitet učenja u odnosu na ResNet-10, ali istovremeno može biti podložniji prenaučivosti kad je dostupan mali broj podataka. Projekt MedicalNet pruža unaprijed naučene 3D ResNet modele namijenjene prijenosnom učenju u zadacima medicinske analize slike, pri čemu su modeli prethodno učeni na skupovima podataka različitih organa i patologija. Isječak koda koji pokazuje pripremu oba ResNet modela vidljiv je na slici 7.9.: funkcija ovisno o odabranoj varijanti inicijalizira ResNet10 ili ResNet18 prilagođen broju ulaznih kanala i klasa. Zatim učitava odgovarajuće prednaučene težine MedicalNet modela. Time se prethodno naučeni dio mreže koristi kao početna točka za

daljnje učenje na ciljnom klasifikacijskom zadatku.

```
def build_and_load_medicalnet(variant, in_channels, n_classes):
    # Build the MedicalNet-compatible 3D ResNet
    if variant == "resnet10":
        model = ResNet3D(
            BasicBlock, layers=[1, 1, 1, 1],
            in_channels, n_classes,
            shortcut_type="B"
        )
    else: # resnet18
        model = ResNet3D(
            BasicBlock, layers=[2, 2, 2, 2],
            in_channels, n_classes,
            shortcut_type="A"
        )

    # Load pretrained MedicalNet weights
    checkpoint = torch.load(
        MEDICALNET_WEIGHTS[variant],
        map_location="cpu"
    )
    state = checkpoint.get("state_dict", checkpoint)
    state = {k.replace("module.", ""): v for k, v in state.items()}

    # New classification head is not in the pretrained checkpoint
    model.load_state_dict(state, strict=False)

    return model
```

Slika 7.9. Isječak koda za izgradnju 3D ResNet arhitekture i učitavanje prethodno naučenih MedicalNet težina.

Kao treća arhitektura korišten je *BrainIAC* (engl. *Brain Imaging Adaptive Core*), model razvijen za analizu strukturnih MR slika mozga i preuzet iz repozitorija na *Hugging Faceu* [32]. Model je prednaučen kontrastivnim učenjem SimCLR. Kontrastivno učenje je tehnika nenadziranog strojnog učenja kojime model uči smjestiti reprezentacije podataka u prostor na takav način da slični podaci u njemu budu bliski, a različiti primjeri udaljeni. Za razliku od ResNet arhitekture, BrainIAC nije temeljen na konvoluciji, već na transformerskoj arhitekturi. Ključni dio transformera je mehanizam pažnje kojim se modeliraju odnosi između različitih dijelova ulaza – Vision Transformer ulazni volumen dijeli na manje trodimenzionalne blokove koji se zatim predstavljaju vektorima. Postavljanje prednaučenih parametara gotovo je identično postupku za ResNet modele na slici 7.9.

Osim pojedinačnih modela, isproban je i pristup ansambla, gdje se predikcija dobiva uprosječavanjem dva modela – onog temeljenog na sivoj i na bijeloj tvari.

Od hiperparametara, ispitan je utjecaj z-score standardizacije (svaki primjer zasebno) te utjecaj načina učenja – zamrznuta ili odmrznuta osnovna mreža (engl. *backbone*). Za BrainIAC model razmatrana je samo konfiguracija sa standardizacijom, što je zadano od strane autora. Tako se kombiniraju tri arhitekture (ResNet-10, ResNet-18 i BrainIAC), dva načina učenja (zamrznuta i odmrznuta osnovna mreža) te dvije varijante predobrade (sa i bez standardizacije). Stopa učenja odabire se ovisno o tome je li osnovna mreža zamrznuta: za zamrznutu se koristi veća stopa učenja (10^{-3} umjesto 10^{-4}).

Dodatno, kao eksperimentalni hiperparametar uključena je i augmentacija ulaznih podataka. Augmentacija je provedena samo na podacima za učenje, pri čemu su za svaki izvorni volumen nasumično generirane dodatne varijante primjenom različitih transformacija. Korištene su četiri vrste transformacija:

1. nasumični pomak – volumen se translacija za slučajno odabranu vrijednost duž svake od triju prostornih osi, pri čemu je maksimalni pomak ograničen na tri voksel.
2. nasumično uvećanje (engl. *zoom*) – volumen se nasumično povećava ili smanjuje unutar raspona od 0,9 do 1,1. Kao i nasumični pomak, simulira manje razlike u položaju.
3. dodavanje Gaussovog šuma – dodaje se šum sa standardnom devijacijom 0,03.
4. elastična deformacija (engl. *elastic deformation*) – volumen se lokalno deformira pomoću glatkih, nasumično generiranih pomaka u svim trima prostornim dimenzijama. Simulira manje anatomske varijacije.

Kod svih prostornih transformacija istovremeno se transformira i maska. Maska binarno određuje koji se vokseli uključuju u daljnju analizu, osiguravajući tako da se analiza provodi nad željenim područjem, umjesto da npr. pozadinski vokseli utječu na klasifikaciju. Maska se računa s pragom od 0,05, odnosno zadržani su oni vokseli koji prosječno po svim ispitanicima imaju jačinu veću od 0,05. Svaka se transformacija primjenjuje neovisno o drugima, čime se za pojedini volumen može generirati više novih primjera.

Isječak koda vidljiv je na slici 7.10.

```
volumes = [volume]

if np.random.rand() < p_translate:
    volumes.append(random_translate(volume, gm_mask, max_shift=max_shift, order=order))
if np.random.rand() < p_zoom:
    volumes.append(random_zoom(volume, gm_mask, zoom_range=zoom_range, order=order))
if np.random.rand() < p_elastic:
    volumes.append(elastic_deform(volume, gm_mask, alpha=elastic_alpha, sigma=elastic_sigma, order=order))
if np.random.rand() < p_noise:
    volumes.append(add_gaussian_noise(volume, gm_mask, noise_std=noise_std))

return volumes
```

Slika 7.10. Isječak koda koji prikazuje augmentaciju podataka.

7.2.3. Objašnjivost

Za oba načina dubokog učenja, dodatno je provedena analiza ponašanja modela putem metode objašnjenja modela Grad-CAM. Grad-CAM mape aktivacija su izračunate na izlazu zadnjeg konvolucijskog sloja te je proveden postupak opisan u poglavlju 4.2.4. i vidljiv na slici 7.11. Grad-CAM mape analizirane su za svaku klasu posebno, pri čemu su se prvo izračunale prosječne VBM mape sudionika kako bi se dobila referentna pozadinska slika preko koje se prikazuju konkretne Grad-CAM mape. Dodatno je provedena analiza točnih i netočnih predikcija. Svaki uzorak svrstan je u jednu od skupina: *true positive* (TP), *true negative* (TN), *false positive* (FP) i *false negative* (FN). Prije izračuna Grad-CAM mapa uzorci su filtrirani prema pouzdanosti predikcije, pri čemu su iz daljnje analize izostavljeni uzorci za koje je vjerojatnost predviđene klase bila manja od zadanog praga. Za preostale uzorke izračunata je Grad-CAM mapa za predviđenu klasu, a u slučaju netočne predikcije dodatno i mapa za stvarnu klasu (slika 7.12.). Time je omogućena usporedba područja na temelju kojih model donosi odluke kod različitih ishoda klasifikacije. Posebno su ispitane usporedbe skupina TP i FN te skupina TP i FP. Za obje usporedbe prikazane su prosječne Grad-CAM mape pojedine skupine te njihove međusobne razlike.

Usporedba skupina TP i FN omogućuje analizu razlika između uspješnog i neuspješnog prepoznavanja bolesti. Na taj se način može ispitati postoje li određena područja koja su izraženije zastupljena kod ispitanika koje je model ispravno klasificirao, a manje zastupljena kod ispitanika kod kojih je pozitivna klasa promašena, odnosno postoje li područja koje model lako uoči te područja koja modelu lako pobjegnu.

```

logits[0, class_idx].backward()

# gradient means give per-channel weights
weights = self._grad.mean(dim=(2, 3, 4), keepdim=True)
cam = F.relu(weights * self._act).sum(dim=1, keepdim=True))

# upsample CAM to original input resolution
cam = F.interpolate(cam, size=x.shape[2:], mode="trilinear", align_corners=False)
cam = cam.squeeze().detach().cpu().numpy()
mn, mx = cam.min(), cam.max()
return (cam - mn) / (mx - mn + 1e-8)

```

Slika 7.11. Isječak koda za izračun Grad-CAM mape aktivacije.

```

for i in range(len(X_t)):
    xi = X_t[i:i+1]
    with torch.no_grad():
        probs = torch.softmax(model(xi), dim=1).squeeze()
        pred = int(probs.argmax().item())
        conf = float(probs[pred].item())
        if conf < conf_min:
            continue
        true = int(true_labels[i])
        cam_pred = gcam.compute(xi, pred)
        cam_true = gcam.compute(xi, true) if true != pred else cam_pred

```

Slika 7.12. Isječak koda za izračun Grad-CAM mapa za predviđenu i stvarnu klasu.

S druge strane, usporedba skupina TP i FP omogućuje analizu razlike između ispravnog i pogrešnog predviđanja bolesti. Ova usporedba posebno je korisna za utvrđivanje koristi li model pri lažno pozitivnim predikcijama slična područja kao pri ispravnim pozitivnim predikcijama. Također pruža uvid u obrasce prisutne kod zdravih pacijenata koje model smatra indikatorima bolesti.

8. Rezultati i rasprava

8.1. Klasično strojno učenje

Prikaz odabranih značajki postupkom pohlepnog pretraživanja unaprijed te postignutih metrika nalazi se u tablicama 8.1.-8.4. U prvom redu u stupcu "Značajke" nalazi se prva odabrana značajka, a u svakom idućem redu navedena je novoodabrana značajka po postupku GFS. Stupci "Točnost" i "F1" prikazuju prosječnu točnost i prosječnu F1 mjeru po pet preklopa vanjske unakrsne validacije, a stupac "Std. devijacija" prikazuje standardnu devijaciju točnosti po preklopima.

Značajke	Točnost	Std. devijacija	F1
VBM sive tvari	0,6353	0,0254	0,6598
WBM bijele tvari	0,5725	0,0147	0,5976
WBM sive tvari	0,5725	0,0493	0,6003
VBM bijele tvari	0,5451	0,0222	0,5706
Volumeni lijeve hemisfere	0,5725	0,0454	0,6015

Tablica 8.1. Rezultati optimizacije hiperparametara za klasifikacijski zadatak kontrola vs BD.

Značajke	Točnost	Std. devijacija	F1
WBM sive tvari	0,7059	0,0333	0,6633
ROI	0,7059	0,0333	0,6633
WBM bijele tvari	0,7176	0,0346	0,6947
VBM sive tvari	0,7176	0,0440	0,6711
Debljina desne hemisfere	0,7176	0,0440	0,6711

Tablica 8.2. Rezultati optimizacije hiperparametara za klasifikacijski zadatak kontrola vs SZ.

Iz prikazanih rezultata vidljivo je da povećavanje broja skupova značajki ne dovodi nužno do poboljšanja klasifikacijskih performansi. U pojedinim slučajevima najbolje su performanse ostvarene već s jednim skupom značajki, dok dodavanje novih značajki nije utjecalo na rezultat ili je dovelo do pogoršanja performansi. Također, dodavanjem zna-

Značajke	Točnost	Std. devijacija	F1
Debljina lijeve hemisfere	0,5469	0,0583	0,5157
Debljina desne hemisfere	0,5628	0,1012	0,5389
Volumen desne hemisfere	0,5704	0,0176	0,5939
WBM sive tvari	0,5706	0,0613	0,4753
VBM bijele tvari	0,5627	0,0368	0,4675

Tablica 8.3. Rezultati optimizacije hiperparametara za klasifikacijski zadatak BD vs SZ.

Značajke	Točnost	Std. devijacija	F1
WBM sive tvari	0,4779	0,0286	0,4733
ROI	0,4779	0,0286	0,4733
Debljina desne hemisfere	0,4544	0,0356	0,4487
Volumen lijeve hemisfere	0,4648	0,0307	0,4594
Volumen desne hemisfere	0,4648	0,0307	0,4594

Tablica 8.4. Rezultati optimizacije hiperparametara za višeklasni klasifikacijski zadatak.

čajki moguće je postići porast točnosti, a pad F1 mjere (npr. prijelaz s tri na četiri skupa značajki kod klasifikacije zdravih i shizofrenih). To pokazuje da dodatne značajke mogu biti redundantne te ne moraju sadržavati dodatne diskriminativne informacije korisne za klasifikaciju, već unijeti šum. Čak i u slučajevima poboljšanja performansi uslijed dodavanja novih skupova značajki, točnost ne postiže značajno poboljšanje (najveći zabilježeni porast je 2.5% za točnost), dok F1 mjera bilježi najveći porast od 6%.

Promatrajući sve klasifikacijske zadatke, posebno se ističu VBM i WBM značajke. VBM značajke sive tvari pokazale su se najvažnijima za razlikovanje zdravih i bipolarnih ispitanika, dok su WBM značajke bile najkorisnije za razlikovanje zdravih i shizofrenih ispitanika te za višeklasni problem.

Odabrani hiperparametri razlikovali su se između preklopa vanjske unakrsne validacije. U klasifikaciji zdravih i bipolarnih prevladava odabir logističke regresije, koja je odabrana u 11 od ukupno 15 konfiguracija (uz poneku slučajnu šumu). Slično je kod klasifikacije bipolarnih i shizofrenih ispitanika, gdje je logistička regresija odabrana u 10 od 15 konfiguracija, najčešće uz jaču regularizaciju. Broj zadržanih značajki kretao se uglavnom između 10 i 100. U preostala dva klasifikacijska zadatka nema konzistentnosti u odabiru konfiguracija.

8.2. Duboko učenje

8.2.1. Učenje od početka

Skupina	Model	Siva tvar		Bijela tvar	
		Točnost	F1	Točnost	F1
1	1 Conv + 1 FC	0,546	0,512	0,462	0,408
	1 Conv + 2 FC	0,555	0,529	0,463	0,384
	2 Conv + 1 FC	0,547	0,505	0,489	0,420
	2 Conv + 2 FC	0,558	0,529	0,477	0,399
2	Normalizacija mini-skupa	0,556	0,538	0,470	0,407
	Normalizacija po grupi	0,554	0,532	0,471	0,400
3	Normalizacija prije aktivacije	0,558	0,539	0,458	0,376
	Normalizacija nakon aktivacije	0,557	0,538	0,460	0,381
4	GAP	0,558	0,540	0,460	0,390
	Flatten	0,574	0,565	0,543	0,531

Tablica 8.5. Prosječne vrijednosti točnosti i F1 mjere za pojedine modele unutar četiriju skupina eksperimenata. Za svaku arhitekturnu konfiguraciju vrijednosti su izračunate kao prosjek rezultata dobivenih na svim četirima klasifikacijskim zadacima (kontrola vs BD, kontrola vs SZ, BD vs SZ, kontrola vs BD vs SZ) i za sve ispitane kombinacije hiperparametara.

U tablici 8.5. prikazani su prosječni rezultati postignuti po skupovima eksperimenata. Navedeno je prikazano kako bi se provjerilo postoje li općeniti trendovi ponašanja modela. Iz prikazanog se nameću dva očita zaključka:

1. VBM sive tvari je puno veće koristi modelu od VBM-a bijele tvari – u svakoj od deset usporedbi model koji koristi VBM sive tvari ima bolje rezultate
2. Korištenje GAP sloja rezultira slabijim rezultatom nego sloja izravnavanja, osobito kod modela s VBM-om bijele tvari. To može biti posljedica očuvanja prostornih informacija. GAP sažima svaku mapu značajki na jednu prosječnu vrijednost, pri čemu se gubi informacija o položaju izdvojenih obilježja, dok Flatten zadržava njihov prostorni raspored.

Ostali eksperimenti ne pokazuju općenite trendove ponašanja, stoga treba prikazati rezultate po konkretnim klasifikacijskim zadacima (tablice 8.6.-8.9.)

Iz dobivenih tablica proizlaze sljedeći zaključci:

1. Kao i prije, najbolje rezultate pokazuje klasifikacija zdravih i shizofrenih osoba, dok ostala tri zadatka ostaju teška. Značajno bolji rezultati mogli bi biti povezani

Skupina	Model	Siva tvar		Bijela tvar	
		Točnost	F1	Točnost	F1
1	1 Conv + 1 FC	0,574	0,535	0,500	0,437
	1 Conv + 2 FC	0,568	0,539	0,509	0,422
	2 Conv + 1 FC	0,565	0,517	0,501	0,399
	2 Conv + 2 FC	0,571	0,539	0,495	0,397
2	Normalizacija mini-skupa	0,577	0,559	0,498	0,417
	Normalizacija po grupi	0,571	0,547	0,510	0,433
3	Normalizacija prije aktivacije	0,575	0,551	0,496	0,398
	Normalizacija nakon aktivacije	0,572	0,544	0,505	0,427
4	GAP	0,581	0,557	0,499	0,413
	Flatten	0,592	0,584	0,548	0,536

Tablica 8.6. Prosječne vrijednosti točnosti i F1 mjere za klasifikacijski zadatak kontrole vs BD. Vrijednosti su izračunate kao prosjek preko svih ispitanih hiperparametarskih konfiguracija.

Skupina	Model	Siva tvar		Bijela tvar	
		Točnost	F1	Točnost	F1
1	1 Conv + 1 FC	0,624	0,594	0,532	0,486
	1 Conv + 2 FC	0,650	0,647	0,533	0,467
	2 Conv + 1 FC	0,641	0,612	0,564	0,514
	2 Conv + 2 FC	0,663	0,657	0,542	0,475
2	Normalizacija mini-skupa	0,648	0,645	0,543	0,499
	Normalizacija po grupi	0,649	0,645	0,546	0,484
3	Normalizacija prije aktivacije	0,652	0,649	0,526	0,465
	Normalizacija nakon aktivacije	0,656	0,654	0,513	0,435
4	GAP	0,649	0,647	0,522	0,465
	Flatten	0,699	0,696	0,642	0,635

Tablica 8.7. Prosječne vrijednosti točnosti i F1 mjere za klasifikacijski zadatak kontrole vs SZ. Vrijednosti su izračunate kao prosjek preko svih ispitanih hiperparametarskih konfiguracija.

Skupina	Model	Siva tvar		Bijela tvar	
		Točnost	F1	Točnost	F1
1	1 Conv + 1 FC	0,549	0,529	0,503	0,448
	1 Conv + 2 FC	0,559	0,547	0,497	0,413
	2 Conv + 1 FC	0,551	0,514	0,540	0,472
	2 Conv + 2 FC	0,556	0,537	0,528	0,453
2	Normalizacija mini-skupa	0,555	0,547	0,507	0,440
	Normalizacija po grupi	0,553	0,536	0,498	0,411
3	Normalizacija prije aktivacije	0,560	0,554	0,500	0,405
	Normalizacija nakon aktivacije	0,560	0,554	0,507	0,407
4	GAP	0,556	0,551	0,502	0,437
	Flatten	0,574	0,565	0,571	0,562

Tablica 8.8. Prosječne vrijednosti točnosti i F1 mjere za klasifikacijski zadatak BD vs SZ. Vrijednosti su izračunate kao prosjek preko svih ispitanih hiperparametarskih konfiguracija.

Skupina	Model	Siva tvar		Bijela tvar	
		Točnost	F1	Točnost	F1
1	1 Conv + 1 FC	0,439	0,389	0,313	0,259
	1 Conv + 2 FC	0,445	0,383	0,313	0,233
	2 Conv + 1 FC	0,432	0,380	0,350	0,293
	2 Conv + 2 FC	0,441	0,384	0,343	0,270
2	Normalizacija mini-skupa	0,444	0,403	0,330	0,273
	Normalizacija po grupi	0,443	0,401	0,327	0,271
3	Normalizacija prije aktivacije	0,443	0,402	0,310	0,236
	Normalizacija nakon aktivacije	0,442	0,400	0,317	0,252
4	GAP	0,446	0,404	0,316	0,246
	Flatten	0,429	0,414	0,409	0,391

Tablica 8.9. Prosječne vrijednosti točnosti i F1 mjere za višeklasni klasifikacijski zadatak. Vrijednosti su izračunate kao prosjek preko svih ispitanih hiperparametarskih konfiguracija.

s izraženijim strukturnim promjenama baš u bijeloj i sivoj tvari povezanim sa shizofrenijom [33].

2. Vrsta normalizacije i njena pozicija nemaju konzistentnih utjecaja, već je njihov utjecaj ovisan o zadatku i arhitekturi. Osim toga, standardna devijacija metrika je značajnija u odnosu na dobitak/gubitak promjenom vrste/pozicije normalizacije (standardna devijacija iznosi oko 0.05-0.1)
3. Za tri od četiri klasifikacijska zadatka (osim klasifikacije zdravih i bipolarnih), uočava se blagi trend prema boljim rezultatima modela s dvije konvolucijske razine pri korištenju VBM-a bijele tvari, dok se kod VBM-a sive tvari nešto boljima pokazuju modeli s dva potpuno povezana sloja. Ipak, razlike između promatranih arhitektura relativno su male, odnosno povećanje složenosti mreže ne dovodi do značajnijeg poboljšanja klasifikacijskih rezultata.

Što se tiče optimizacije hiperparametara, modeli koji koriste VBM sive tvari nisu pokazali naklonjenost određenoj konfiguraciji, dok modeli s VBM-om bijele tvari naginju konfiguracijama s većom jezgrom (trend je izraženiji kod arhitektura s jednim konvolucijskim slojem) i većom regijom sažimanja (trend je izraženiji kod arhitektura s dva konvolucijska sloja), vidljivo u tablicama 8.10. i 8.11. Moguće objašnjenje jest da morfološke promjene bijele tvari ne moraju biti ograničene na vrlo lokalizirana područja, već se mogu očitovati kao prostorno širi obrasci, a veće konvolucijske jezgre omogućavaju obuhvaćanje šireg konteksta. Veće sažimanje najvećom vrijednošću dodatno sažima

prostorne informacije i smanjuje osjetljivost na manje lokalne varijacije. Navedena opa-

Arhitektura	Kernel	Točnost	F1
1 Conv + 1 FC	3	0,450	0,392
	5	0,470	0,423
1 Conv + 2 FC	3	0,461	0,368
	5	0,465	0,400

Tablica 8.10. Utjecaj veličine konvolucijske jezgre na prosječne klasifikacijske rezultate pojedinih arhitektura za podatke bijele tvari.

Arhitektura	Sažimanje	Točnost	F1
2 Conv + 1 FC	[2,2]	0,471	0,384
	[2,4]	0,495	0,412
	[4,4]	0,519	0,493
2 Conv + 2 FC	[2,2]	0,459	0,369
	[2,4]	0,478	0,400
	[4,4]	0,495	0,427

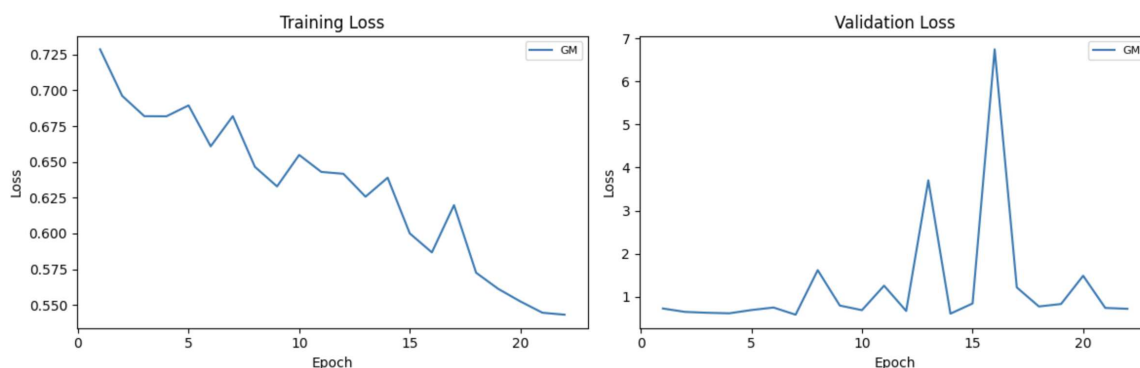
Tablica 8.11. Utjecaj veličine sažimanja najvećom vrijednošću na prosječne klasifikacijske rezultate arhitektura s dvije konvolucijske razine za podatke bijele tvari.

žanja predstavljaju opće trendove uočene analizom prosječnih rezultata kroz različite arhitekture i hiperparametarske konfiguracije. Međutim, ona ne podrazumijevaju da konfiguracija koja prati sve navedene trendove nužno ostvaruje najbolji pojedinačni rezultat za svaki klasifikacijski zadatak. Optimalna kombinacija hiperparametara razlikuje se ovisno o zadatku. Stoga su u tablici 8.12. prikazane konfiguracije koje su se pokazale najboljima (po mjeri točnosti) za svaki od promatranih klasifikacijskih zadataka.

Zadatak	Model	Reprezentacija	Jezgra	Sažimanje	Kanali	Točnost	F1
Kontrola vs BD	Flatten	Siva tvar	5	2	32	$0,624 \pm 0,049$	$0,616 \pm 0,050$
Kontrola vs SZ	Flatten	Siva tvar	5	2	8	$0,737 \pm 0,052$	$0,735 \pm 0,054$
BD vs SZ	2 Conv + 2 FC	Bijela tvar	[5, 3]	[4, 4]	[8, 16, 8]	$0,618 \pm 0,051$	$0,605 \pm 0,060$
Višeklasna	2 Conv + 2 FC	Siva tvar	[5, 5]	[2, 2]	[4, 8, 4]	$0,475 \pm 0,021$	$0,433 \pm 0,050$

Tablica 8.12. Najbolje konfiguracije za pojedine klasifikacijske zadatke prema prosječnoj točnosti. Za svaki je zadatak prikazana konfiguracija s najvećom točnošću između svih eksperimenata provedenih na podacima bijele i sive tvari. Rezultati su prikazani kao prosjek $\pm$ standardna devijacija.

Slika 8.1. pokazuje kako gubitak na skupu za validaciju blago raste, što upućuje na to da je odluka o zaustavljanju učenja bila točna. Vidljiv nagli rast oko petnaeste epohe može se pripisati maloj veličini validacijskog skupa, zbog čega promjene u predikcijama malog broja uzoraka mogu uzrokovati velike oscilacije validacijskog gubitka.



Slika 8.1. Kretanje funkcije gubitka na skupu za učenje i validaciju tijekom učenja modela za klasifikacijski zadatak kontrole vs BD.

Važan zaključak dobiven pri početku odnosi se na potrebu za skaliranjem ulaznih podataka. Pri učenju modela na neskaliranim podacima uočeno je degenerirano ponašanje modela, pri kojem su svi ulazni primjeri klasificirani u samo jednu klasu. Takvo se ponašanje ponavljalo neovisno o ispitanoj arhitekturi, zbog čega rezultati tih eksperimenata nisu uključeni u daljnju usporedbu. Nakon uvođenja skaliranja ulaznih podataka modeli su počeli razlikovati klase – skaliranje se stoga pokazalo kao nužan korak predobrade podataka.

8.2.2. Prednaučene mreže

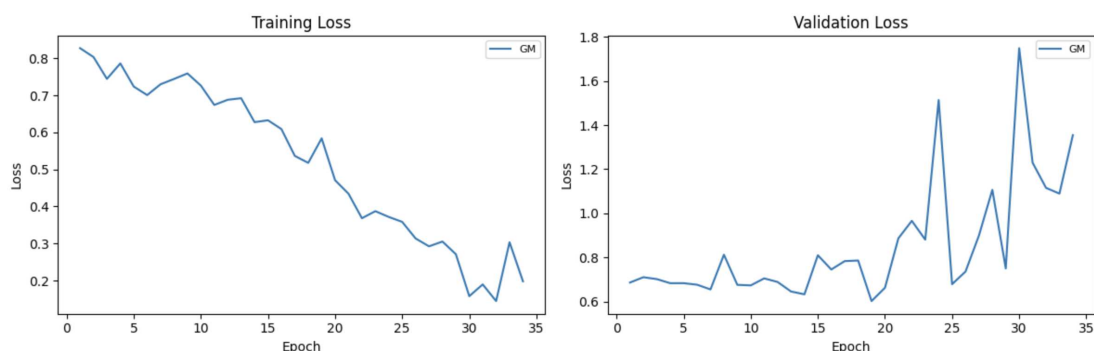
Na temelju provedenih eksperimenata i usporedbe metrika u tablicama 8.13. i 8.14., moguće je izdvojiti nekoliko ključnih opažanja:

- U skladu s prijašnjim rezultatima, modeli sa sivom tvari daju bolje rezultate od modela s bijelom tvari te i dalje najbolje rezultate pokazuje klasifikacija zdravih i shizofrenih osoba.
- Modeli koji uče i osnovni dio prednaučene mreže (odmrznuta osnovna mreža) ostvaruju značajno bolje rezultate u gotovo svim promatranim slučajevima – svi zadaci, osim klasifikacije zdravih i bipolarnih, bilježe velik rast i točnosti i F1 mjere. Odmrzavanje dijela mreže omogućuje prilagodbu težina ovom specifičnom problemu, u suprotnom je mreža ograničena na reprezentacije naučene tijekom predučenja, koje ne moraju biti optimalne za ovaj klasifikacijski problem.
- Augmentacija podataka ne doprinosi konzistentnom poboljšanju performansi. Nje-

zin učinak ovisi o klasifikacijskom zadatku i korištenim podacima. Dok se kod zadatka klasifikacije zdravih i bipolarnih uočava blago poboljšanje, kod ostalih zadataka rezultati su mješoviti ili nešto lošiji, zbog čega se ne može utvrditi opća prednost korištene augmentacije. Moguće je da primijenjene transformacije ne stvaraju dovoljno novih diskriminativnih informacija.

- Za modele sa sivom tvari standardno skaliranje dovodi do smanjenja točnosti i F1 mjere u svim promatranim klasifikacijskim zadacima. Takav obrazac nije prisutan kod modela s bijelom tvari, kod kojih su razlike znatno manje i često idu u korist standardnog skaliranja.
- Ne uočava se jasna korist kombiniranja modela bijele i sive tvari u ansambl – ako modeli ne čine dovoljno različite pogreške, njihovim kombiniranjem ne mora doći do poboljšanja konačne predikcije.
- Najboljom arhitekturom se generalno pokazao ResNet18, ali ResNet10 često ima slične rezultate, dok se čak za klasifikaciju zdravih i bipolarnih pokazao najboljim. S druge strane, BrainIAC model pokazao se najlošijim. To nije iznenađujuće, budući da BrainIAC model ima najviše parametara, stoga možda zahtijeva veći broj ispitanika.

Kao i kod učenja mreže od početka, najbolji rezultati ne prate nužno gore objašnjene općenite trendove. U tablici 8.15. prikazane su konfiguracije koje su postigle najveće točnosti za pojedini klasifikacijski zadatak.



Slika 8.2. Kretanje funkcije gubitka na skupu za učenje i validaciju tijekom učenja modela BD vs. SZ.

Na slici 8.2. ponovno se vidi isti trend kao i kod učenja mreže od početka – gubitak

Zadatak	Postavka	Siva tvar		Bijela tvar		Ansambl	
		Točnost	F1	Točnost	F1	Točnost	F1
Zamrznuta osnovna mreža							
Kontrola vs BD	Zamrznuta	0,512	0,379	0,501	0,339	0,507	0,359
	Odmrznuta	0,539	0,465	0,500	0,396	0,535	0,457
Kontrola vs SZ	Zamrznuta	0,544	0,438	0,502	0,334	0,535	0,413
	Odmrznuta	0,693	0,680	0,601	0,577	0,697	0,691
BD vs SZ	Zamrznuta	0,513	0,390	0,500	0,337	0,511	0,367
	Odmrznuta	0,580	0,549	0,561	0,526	0,574	0,551
Višeklasna	Zamrznuta	0,359	0,248	0,331	0,212	0,348	0,242
	Odmrznuta	0,446	0,414	0,375	0,313	0,453	0,425
Augmentacija							
Kontrola vs BD	Bez aug.	0,518	0,406	0,495	0,358	0,514	0,397
	S aug.	0,533	0,438	0,507	0,377	0,527	0,419
Kontrola vs SZ	Bez aug.	0,615	0,552	0,552	0,457	0,619	0,555
	S aug.	0,622	0,566	0,551	0,454	0,613	0,549
BD vs SZ	Bez aug.	0,550	0,476	0,537	0,445	0,545	0,461
	S aug.	0,542	0,463	0,523	0,418	0,540	0,457
Višeklasna	Bez aug.	0,406	0,334	0,348	0,253	0,401	0,337
	S aug.	0,399	0,328	0,358	0,272	0,401	0,330
Standardno skaliranje							
Kontrola vs BD	Bez skaliranja	0,544	0,469	0,499	0,372	0,525	0,428
	Sa skaliranjem	0,513	0,398	0,499	0,358	0,519	0,398
Kontrola vs SZ	Bez skaliranja	0,670	0,658	0,543	0,444	0,652	0,625
	Sa skaliranjem	0,583	0,493	0,564	0,473	0,599	0,513
BD vs SZ	Bez skaliranja	0,567	0,530	0,529	0,423	0,557	0,490
	Sa skaliranjem	0,526	0,415	0,535	0,443	0,537	0,441
Višeklasna	Bez skaliranja	0,439	0,393	0,350	0,262	0,426	0,376
	Sa skaliranjem	0,381	0,294	0,361	0,277	0,389	0,319

Tablica 8.13. Prosječne vrijednosti točnosti i F1 mjere za različite hiperparametarske postavke i klasifikacijske zadatke. Vrijednosti su izračunate kao prosjek preko svih arhitektura (Resnet10, Resnet18, BrainIAC) i ostalih hiperparametarskih konfiguracija. Primjerice, rezultati za zamrznutu osnovnu mrežu obuhvaćaju sve konfiguracije sa zamrznutom osnovnom mrežom, neovisno o arhitekturi te primjeni augmentacije i standardnog skaliranja. Za analizu standardnog skaliranja uključene su arhitekture ResNet10 i ResNet18 zbog dostupnosti usporedivih konfiguracija sa standardnim skaliranjem i bez njega.

Zadatak	Brainiac	ResNet10	ResNet18
Kontrola vs BD	0,440	0,483	0,439
Kontrola vs SZ	0,634	0,701	0,709
BD vs SZ	0,533	0,537	0,573
Višeklasna	0,388	0,411	0,458
Prosjek	0,499	0,533	0,545

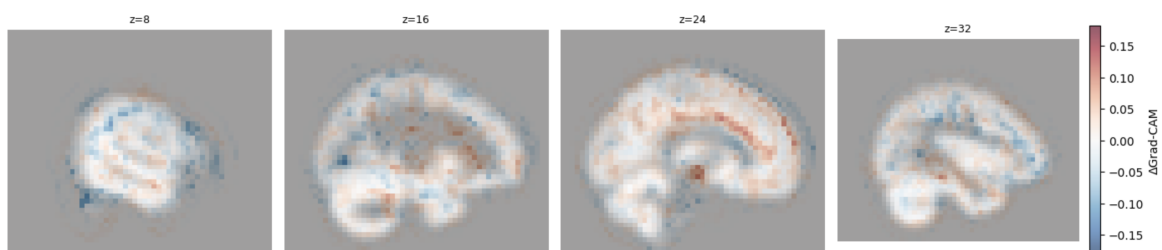
Tablica 8.14. Prosječne vrijednosti F1 mjere nezamrznutih modela prema arhitekturi i klasifikacijskom zadatku. Vrijednosti su izračunate kao prosjek preko odgovarajućih hiperparametarskih konfiguracija.

Zadatak	Model	Reprezentacija	Zamrznuta osnovna mreža	Stand. skaliranje	Augmentacija	Točnost	F1
Kontrola vs BD	ResNet10	Siva tvar	Ne	Ne	Da	0,592 ± 0,078	0,563 ± 0,120
Kontrola vs SZ	ResNet18	Ensemble	Ne	Ne	Da	0,749 ± 0,070	0,748 ± 0,071
BD vs SZ	Brainiac	Siva tvar	Ne	Da	Da	0,629 ± 0,053	0,626 ± 0,054
Višeklasna	ResNet18	Ensemble	Ne	Da	Ne	0,491 ± 0,026	0,482 ± 0,032

Tablica 8.15. Najbolje konfiguracije hiperparametara za pojedine klasifikacijske zadatke prema prosječnoj točnosti unakrsne validacije. Rezultati su prikazani kao prosjek $\pm$ standardna devijacija.

na validacijskom skupu raste dok se na skupu za učenje smanjuje.

Daljnjom analizom najboljih konfiguracija pomoću metode objašnjenja modela Grad-CAM izdvojeni su primarno primjeri za klasifikaciju kontrole vs BD i kontrole vs SZ i to s prednaučenim mrežama. Njima je posvećeno više pažnje jer su rezultirali informativnijim mapama, dok je za ostale kombinacije teže izvući zaključke. Primjerice, slika 8.3. prikazuje razliku između Grad-CAM aktivacija kod TP i FN primjera kod klasifikacijskog zadatka kontrole vs BD. s modelom naučenim ispočetka. Razlike su raspršene po cijeloj slici, umjesto da su lokalizirane – ne otkrivaju važne informacije. Na slici 8.4. ponovno je prikazan loš primjer – pokazuje prosječne Grad-CAM aktivacije za primjere iz bipolarne skupine. Na slici nije moguće uočiti važne regije jer su aktivacije vidljive na cijeloj slici. Lošije i manje interpretabilne Grad-CAM mape u navedenim zadacima mogu upućivati na to da model nije uspio izdvojiti jasne značajke, ponovno ukazujući na problem razlikovanja dvije bolesne osobe. Bolja performansa prednaučenih mreža može biti povezana s dubinom mreža i hijerarhijskom organizacijom značajki – prvi konvolucijski slojevi uglavnom izdvajaju jednostavne značajke, poput rubova, a dublji slojevi to kombiniraju u prostorno šire, jasnije obrasce.



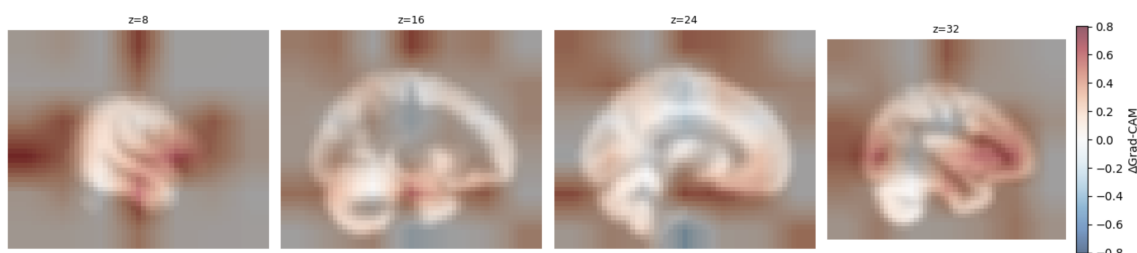
Slika 8.3. Razlika između Grad-CAM aktivacija između TP i FN primjera kod klasifikacijskog zadatka kontrole vs BD.

Na slikama 8.5. i 8.6. prikazane su razlike između prosječnih Grad-CAM aktivacija za TP i FN, odnosno TP i FP primjere kod klasifikacijskog zadatka kontrole vs BD mrežom ResNet10. Crvenija područja pokazuju regije kod kojih je veća aktivacija postignuta u

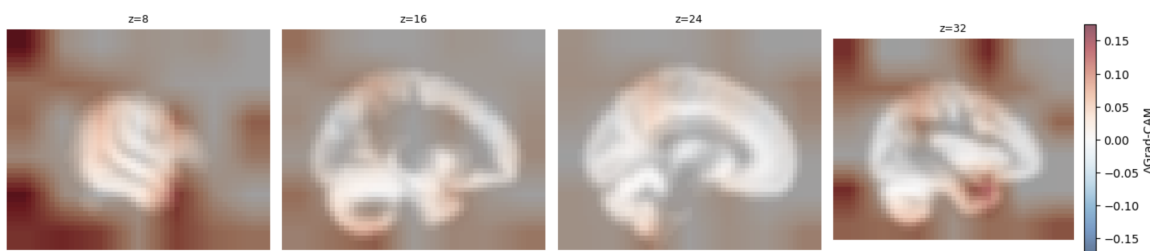


Slika 8.4. Prosječne Grad-CAM aktivacije za primjere iz bipolarne skupine kod višeklasne klasifikacije mrežom ResNet18.

TP primjerima, a kod plavih područja su to FN, odnosno FP primjeri. U oba su slučaja pokazane jače aktivacije u gornjem središnjem dijelu mozga kod netočno klasificiranih primjera. Navedeno može upućivati na to da se model pri donošenju pogrešnih odluka u većoj mjeri oslanja na značajke iz tog područja, a moguće je da te značajke nisu relevantne za razlikovanje skupina kontrole i BD. S druge strane, aktivacije za točno klasificirane primjere su vidljivije u donjem desnom dijelu mozga. Vrijedi spomenuti i činjenicu da su razlike izraženije kod razlikovanja TP i FN primjera – skala jačine se kreće $(-0.8, 0.8)$, u odnosu na $(-0.15, 0.15)$.



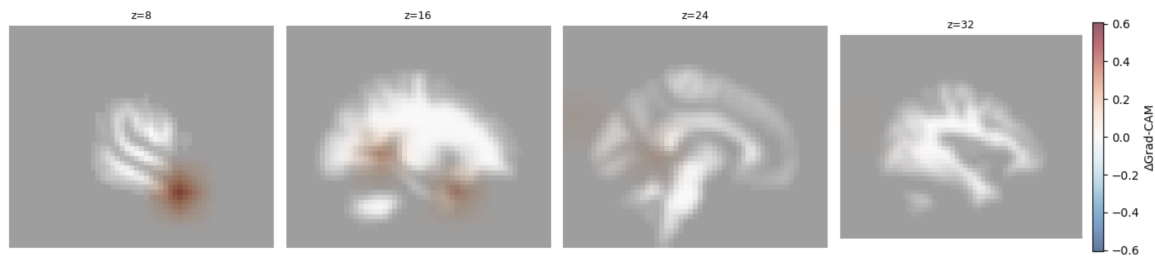
Slika 8.5. Razlika između Grad-CAM aktivacija između TP i FN primjera kod klasifikacijskog zadatka kontrole vs BD mrežom ResNet10



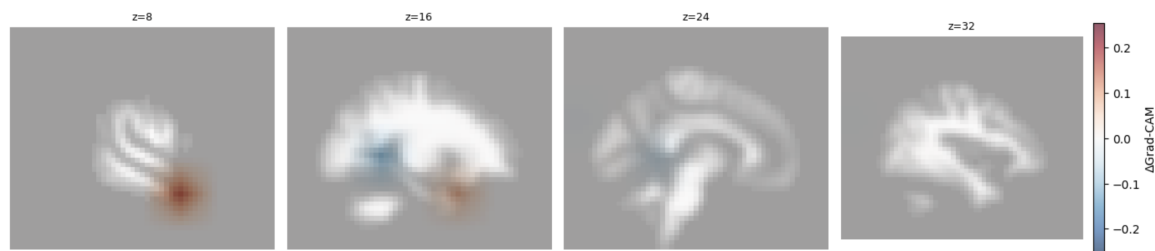
Slika 8.6. Razlika između Grad-CAM aktivacija između TP i FP primjera kod klasifikacijskog zadatka kontrole vs BD mrežom ResNet10

Zanimljiv obrazac uočava se usporedbom TP–FN i TP–FP mapa kod klasifikacijskog zadatka kontrole vs SZ. s bijelom tvari, slike 8.7. i 8.8., pri čemu se na pojedinim lokacijama pojavljuju razlike suprotnog predznaka. Pozitivna razlika između TP i FN primjera

pokazuje veću Grad-CAM atribuciju kod TP primjera, dok negativna razlika između TP i FP primjera na istom području pokazuje veću atribuciju kod FP primjera. Takav obrazac može upućivati na povezanost značajki iz tog područja s predviđanjem SZ klase, pri čemu njihova slabija zastupljenost prati propuštene SZ primjere, dok se njihova izražena zastupljenost pojavljuje i kod kontrolnih primjera pogrešno klasificiranih kao SZ.



Slika 8.7. Razlika između Grad-CAM aktivacija između TP i FN primjera kod klasifikacijskog zadatka kontrole vs SZ. mrežom ResNet18 i slikama bijele tvari.



Slika 8.8. Razlika između Grad-CAM aktivacija između TP i FP primjera kod klasifikacijskog zadatka kontrole vs SZ. mrežom ResNet18 i slikama bijele tvari.

Zaključno, na mnogim Grad-CAM mapama, a i ovim prikazanim, uočene su izražene atribucije izvan područja mozga. To upućuje na mogućnost da se modeli oslanjaju na nerelevantne značajke kod klasifikacije – to također može biti uzrok nezadovoljavajućim rezultatima.

8.3. Usporedba s postojećim istraživanjima

Dobiveni rezultati u velikoj su mjeri u skladu s rezultatima prethodnih istraživanja opisanih u poglavlju 5. Posebno je relevantna usporedba s radom [23], budući da je u njemu korišten isti skup podataka te slične vrste značajki i klasifikacijskih metoda. U navedenom radu najbolje točnosti iznosile su 63% za razlikovanje zdravih i bipolarnih ispitanika, 75% za razlikovanje zdravih i shizofrenih ispitanika te 62% za razlikovanje bi-

polarnih i shizofrenih ispitanika. Pritom se pokazalo da na rezultate više utječe odabir značajki nego samog klasifikacijskog algoritma. Sličan obrazac vidljiv je i u ovom istraživanju. Primjenom klasičnog strojnog učenja ostvarene su točnosti od 63,5% za klasifikaciju zdravih i bipolarnih ispitanika, 71,8% za zdrave i shizofrene te 57,1% za bipolarne i shizofrene ispitanike. Rezultati su stoga vrlo bliski onima iz rada [23], osobito za klasifikaciju zdravih i bipolarnih ispitanika. Nešto slabiji rezultat dobiven je za razlikovanje bipolarnih i shizofrenih ispitanika, dok se u oba istraživanja najuspješnijim pokazalo razlikovanje zdravih i shizofrenih osoba. Takav rezultat dodatno potvrđuje da su razlike između dviju skupina pacijenata teže prepoznatljive iz korištenih strukturnih MR značajki nego razlike između zdravih ispitanika i osoba sa shizofrenijom.

Sličnost s navedenim radom vidljiva je i u važnosti odabira reprezentacije podataka. U ovom su istraživanju VBM značajke sive tvari bile najkorisnije za razlikovanje zdravih i bipolarnih ispitanika, dok su se WBM značajke posebno istaknule kod razlikovanja zdravih i shizofrenih ispitanika. Nadalje, dodavanje većeg broja skupova značajki nije konzistentno poboljšavalo rezultate. Time se potvrđuje opažanje iz rada [23], u kojem kombiniranje većeg broja značajki uz redukciju dimenzionalnosti također nije dovelo do boljih klasifikacijskih performansi.

Dobiveni rezultati mogu se usporediti i s ostalim istraživanjima iz poglavlja 5., uz oprez zbog razlika u skupovima podataka i metodologiji. U radu [22] razlikovanje bipolarnih i zdravih ispitanika ostvarilo je točnost od 65,23%, što je usporedivo s najboljim rezultatima dobivenima u ovom istraživanju. S druge strane, rad [24], koji koristi funkcionalne MR podatke, za isti binarni zadatak postiže 71,6%. Viša točnost može biti povezana s korištenjem funkcionalnih vremenskih informacija koje nisu prisutne u strukturnim podacima korištenima u ovom radu. Međutim, četveroklasna klasifikacija u istom istraživanju ostvarila je samo 46%, što ponovno pokazuje koliko je istovremeno razlikovanje više psihijatrijskih poremećaja težak problem.

Sveukupno, usporedba s postojećim istraživanjima pokazuje da dobiveni rezultati slijeđe slične trendove kao prethodni radovi. Najpouzdanije se razlikuju zdravi ispitanici i osobe sa shizofrenijom, dok razlikovanje bipolarnog poremećaja od shizofrenije te višeklasna klasifikacija ostaju zahtjevniji problemi. Istovremeno, rezultati ne pokazuju da povećanje složenosti modela samo po sebi dovodi do boljih performansi. Veći utjecaj

imaju odabir odgovarajuće reprezentacije podataka, količina dostupnih podataka i sam klasifikacijski zadatak. To je posebno važno u kontekstu ograničenog skupa podataka korištenog u ovom istraživanju te upućuje na potrebu za većim i heterogenijim skupovima podataka u budućim istraživanjima.

9. Zaključak

Cilj ovog diplomskog rada bio je istražiti mogućnost primjene strojnog i dubokog učenja za automatsko dijagnosticiranje pacijenata s bipolarnim ili shizofrenim poremećajem pomoću podataka magnetske rezonancije mozga. Razvoj ovih metoda od velike je važnosti jer se spomenuti poremećaji djelomično preklapaju u simptomima, što otežava dijagnozu. Navedene metode bi mogle služiti liječnicima i omogućiti raniji početak liječenja.

Rezultati dobiveni primjenom klasičnog strojnog učenja pokazali su da veći broj značajki i veća složenost ne rezultiraju nužno boljim performansama. Važnijim se pokazao odabir odgovarajućih značajki – VBM i WBM značajke su dominirale. Osim toga, logistička regresija često je odabrana kao najbolji klasifikacijski model, što ponovno naglašava važnost kombinacije hiperparametara, umjesto oslanjanja na povećanje složenosti. Slično se moglo zaključiti i u eksperimentima s dubokim učenjem, gdje se povećanje složenosti pokazalo neisplativim. Ni primjena prijenosnog učenja nije donijela znatno poboljšanje u odnosu na ostale pristupe. U tim eksperimentima se ResNet18 općenito pokazao najboljim modelom, a BrainIAC najlošijim. To dodatno potvrđuje zaključak o složenosti, budući da BrainIAC ima najviše parametara.

Glavno ograničenje ovog istraživanja predstavlja malen skup dostupnih podataka, osobito u odnosu na njihovu dimenzionalnost. To može biti uzrok opažanju da duboko učenje nije ostvarilo poboljšanje u odnosu na klasično strojno učenje. Uz to, potrebno je razmotriti mogućnost da podaci magnetske rezonancije mozga u ovom obliku nemaju dovoljnu diskriminativnu moć za pouzdano razlikovanje promatranih skupina. Buduća istraživanja bi zato mogla imati koristi od prikupljanja više podataka, ali i istraživanja dodatnih kliničkih mogućnosti.

Primjena metode objašnjenja modela Grad-CAM ukazala je na važnost objašnjivosti modela u medicinskom području. Kod pojedinih zadataka uočene su izraženije aktivacije u područjima izvan mozga, što upućuje na to da se modeli znaju oslanjati na nevažne informacije pri odlučivanju. Stoga sama analiza točnosti, F1 mjere i sličnih pokazatelja nije dovoljna, već je potrebno razumjeti na temelju čega je model donio odluku.

Zaključno, rezultati ovog rada ukazuju na potencijal primjene metoda strojnog i dubokog učenja u analizi *neuroimaging* podataka, ali istodobno ističu njihova trenutačna ograničenja. Daljnji napredak ovisit će prvenstveno o dostupnosti podataka, ali i o pažljivom odabiru značajki, modela i njihovih hiperparametara.

Literatura

- [1] J. D. Lish, S. Dime-Meenan, P. C. Whybrow, R. A. Price, i R. M. A. Hirschfeld, “The national depressive and manic-depressive association (dmda) survey of bipolar members”, *Journal of Affective Disorders*, sv. 31, br. 4, str. 281–294, kolovoz 1994. [https://doi.org/10.1016/0165-0327\(94\)90104-X](https://doi.org/10.1016/0165-0327(94)90104-X)
- [2] A. Bradford, A. N. D. Meyer, S. Khan, T. D. Giardina, i H. Singh, “Diagnostic error in mental health: A review”, *Diagnosis*, sv. 11, br. 3, str. 189–198, srpanj 2024. <https://doi.org/10.1515/dx-2023-0115>
- [3] A. J. Ferrari, A. J. Baxter, i H. A. Whiteford, “A systematic review of the global distribution and availability of prevalence data for bipolar disorder”, *Journal of Affective Disorders*, sv. 134, br. 1–3, str. 1–13, studeni 2011. <https://doi.org/10.1016/j.jad.2010.11.007>
- [4] W. V. Bobo, “The diagnosis and management of bipolar i and ii disorders: Clinical practice update”, *Mayo Clinic Proceedings*, sv. 92, br. 10, str. 1532–1551, studeni 2017., review article. <https://doi.org/10.1016/j.mayocp.2017.06.022>
- [5] Osmosis, “Graphical representation of bipolar disorder and cyclothymia”, https://commons.wikimedia.org/wiki/File:Bipolar_mood_shifts.png, 2016., screenshot from the video “Bipolar Disorder” at 3:30 and 3:51. Wikimedia Commons. Accessed: 2025-04-20.
- [6] M. J. McCarthy, M. J. L. L. Roux, H. Wei, S. Beesley, J. R. Kelsoe, i D. K. Welsh, “Calcium channel genes associated with bipolar disorder modulate lithium’s amplification of circadian rhythms”, *Neuropharmacology*, sv. 101, str. 439–448, veljača 2016., epub 2015 Oct 22. <https://doi.org/10.1016/j.neuropharm.2015.10.017>

- [7] A. Mardon, *Bipolar Disorder*. Edmonton, AB: Golden Meteorite Press, 2021.
- [8] Wikipedia contributors, "Schizophrenia", <https://en.wikipedia.org/wiki/Schizophrenia>, 2026., wikipedia, The Free Encyclopedia. Accessed: 2026-08-01.
- [9] Schizophrenia Working Group of the Psychiatric Genomics Consortium, "Biological insights from 108 schizophrenia-associated genetic loci", *Nature*, sv. 511, br. 7510, str. 421–427, kolovoz 2014. <https://doi.org/10.1038/nature13595>
- [10] National Institute of Standards and Technology, "New measurements aim to advance and validate portable mri technology", <https://www.technologynetworks.com/analysis/news/new-measurements-aim-to-advance-and-validate-portable-mri-technology-376328>, 2023., [mrežno; stranica posjećena: travanj 2025.].
- [11] K. Maher, "T1 weighted, t2 weighted and pd weighted mri scans", <https://commons.wikimedia.org/wiki/File:T1t2PD.jpg>, 2006., [slika u javnom vlasništvu; prilagođena za potrebe rada; mrežno; stranica posjećena: travanj 2025.].
- [12] Wikipedia contributors, "Machine learning", https://en.wikipedia.org/wiki/Machine_learning, lipanj 2024., [mrežno; stranica posjećena: 3.6.2024.].
- [13] I. H. Witten, E. Frank, M. A. Hall, i C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4. izd. Morgan Kaufmann, listopad 2016.
- [14] Manning Publications, "Sigmoid function", <https://livebook.manning.com/concept/machine-learning/sigmoid-function>, [mrežno; stranica posjećena: 1.8.2026.].
- [15] J. Šnajder, "Su1-2022-p15-bayesovklasifikator", https://www.fer.unizg.hr/_download/repository/SU1-2022-P15-BayesovKlasifikator.pdf, 2022., [mrežno; stranica posjećena: 1.8.2026.].
- [16] Nicoguardo, "Gaussianscatterpca.svg", <https://commons.wikimedia.org/wiki/File:GaussianScatterPCA.svg>, 2008., wikimedia Commons, pristupljeno: 2. kolovoza 2026.

- [17] GeeksforGeeks. (2025) What is perceptron? Pristupljeno: 2. kolovoza 2026. [Mrežno]. Adresa: <https://www.geeksforgeeks.org/machine-learning/what-is-perceptron-the-simplest-artificial-neural-network/>
- [18] V. Dumoulin i F. Visin, “Convolution arithmetic – no padding strides”, https://commons.wikimedia.org/wiki/File:Convolution_arithmetic_-_No_padding_strides.gif, 2016., animation on Wikimedia Commons. Source: https://github.com/vdumoulin/conv_arithmetic. Accessed: 2026-08-07.
- [19] O. Dreessen, “Training convolutional neural networks: What is machine learning?—part 2”, *Analog Dialogue*, sv. 57, br. 1, 2023. [Mrežno]. Adresa: <https://www.analog.com/en/resources/analog-dialogue/articles/training-convolutional-neural-networks-what-is-machine-learning-part-2.html>
- [20] NPR, “Facebook apologizes after ai labels black men as primates”, *NPR*, rujan 2021. [Mrežno]. Adresa: <https://www.npr.org/2021/09/04/1034368231/facebook-apologizes-ai-labels-black-men-primates-racial-bias>
- [21] A. J. Aiya, N. Wani, M. Ramani *et al.*, “Optimized deep learning for brain tumor detection: a hybrid approach with attention mechanisms and clinical explainability”, *Scientific Reports*, sv. 15, str. 31386, kolovoz 2025. <https://doi.org/10.1038/s41598-025-04591-3>
- [22] A. Nunes, H. G. Schnack, C. R. K. Ching, I. Agartz, T. N. Akudjedu, M. Alda, D. Alnæs, S. Alonso-Lana, J. Bauer, B. T. Baune, E. Bøen, C. d. M. Bonnin, G. F. Busatto, E. J. Canales-Rodríguez, D. M. Cannon, X. Caseras, T. M. Chaim-Avancini, U. Dannlowski, A. M. Díaz-Zuluaga, B. Dietsche, N. T. Doan, E. Duchesnay, T. Elvåshagen, D. Emden, L. T. Eyler, M. Fatjó-Vilas, P. Favre, S. F. Foley, J. M. Fullerton, D. C. Glahn, J. M. Goikolea, D. Grotegerd, T. Hahn, C. Henry, D. P. Hibar, J. Houenou, F. M. Howells, N. Jahanshad, T. Kaufmann, J. Kenney, T. T. J. Kircher, A. Krug, T. V. Lagerberg, R. K. Lenroot, C. López-Jaramillo, R. Machado-Vieira, U. F. Malt, C. McDonald, P. B. Mitchell, B. Mwangi, L. Nabulsi, N. Opel, B. J. Overs, J. A. Pineda-Zapata, E. Pomarol-Clotet, R. Redlich, G. Roberts, P. G. Rosa, R. Salvador, T. D. Satterthwaite, J. C. Soares, D. J. Stein, H. S. Temmingh, T. Trappenberg, A. Uhlmann, N. E. M. van Haren, E. Vieta,

- L. T. Westlye, D. H. Wolf, D. Yüksel, M. V. Zanetti, O. A. Andreassen, P. M. Thompson, i T. Hajek, “Using structural mri to identify bipolar disorders – 13 site machine learning study in 3020 individuals from the enigma bipolar disorders working group”, *Molecular Psychiatry*, sv. 25, br. 9, str. 2130–2143, 2020. <https://doi.org/10.1038/s41380-018-0228-9>
- [23] R. Salvador, J. Radua, E. J. Canales-Rodríguez, A. Solanes, S. Sarró, J. M. Goikolea, A. Valiente, G. C. Monté, M. d. C. Natividad, A. Guerrero-Pedraza, N. Moro, P. Fernández-Corcuera, B. L. Amann, T. Maristany, E. Vieta, P. J. McKenna, i E. Pomarol-Clotet, “Evaluation of machine learning algorithms and structural features for optimal mri-based diagnostic prediction in psychosis”, *PLOS ONE*, sv. 12, br. 4, str. e0175683, 2017. <https://doi.org/10.1371/journal.pone.0175683>
- [24] W. Yan, M. Zhao, Z. Fu, G. D. Pearlson, J. Sui, i V. D. Calhoun, “Mapping relationships among schizophrenia, bipolar and schizoaffective disorders: A deep classification and clustering framework using fmri time series”, *Schizophrenia Research*, sv. 245, str. 141–150, 2022. <https://doi.org/10.1016/j.schres.2021.02.007>
- [25] J. Oh, B.-L. Oh, K.-U. Lee, J.-H. Chae, i K. Yun, “Identifying schizophrenia using structural mri with a deep learning algorithm”, *Frontiers in Psychiatry*, sv. 11, str. 16, 2020. <https://doi.org/10.3389/fpsy.2020.00016>
- [26] R. Salvador, J. Radua, E. Canales-Rodríguez, A. Solanes, S. Sarró, J. M. Goikolea, A. Valiente, G. C. Monté, M. d. C. Natividad, A. Guerrero-Pedraza, N. Moro, P. Fernández-Corcuera, B. L. Amann, T. Maristany, E. Vieta, P. J. McKenna, i E. Pomarol-Clotet, “Evaluation of machine learning algorithms and structural features for optimal mri-based diagnostic prediction in psychosis”, 2017. <https://doi.org/10.5281/zenodo.460878>
- [27] E. J. Canales-Rodríguez, J. Radua, E. Pomarol-Clotet, S. Sarró, Y. Alemán-Gómez, Y. Iturria-Medina, i R. Salvador, “Statistical analysis of brain tissue images in the wavelet domain: Wavelet-based morphometry”, *NeuroImage*, sv. 72, str. 214–226, 2013. <https://doi.org/10.1016/j.neuroimage.2013.01.058>
- [28] A. Antoniadis, I. Gijbels, i G. Grégoire, “Model selection using wavelet

decomposition and applications”, *Biometrika*, sv. 84, br. 4, str. 751–763, 1997.
<https://doi.org/10.1093/biomet/84.4.751>

- [29] P. Probst, M. N. Wright, i A.-L. Boulesteix, “Hyperparameters and tuning strategies for random forest”, *WIREs Data Mining and Knowledge Discovery*, sv. 9, br. 3, str. e1301, 2019. <https://doi.org/https://doi.org/10.1002/widm.1301>
- [30] D. Kim, W. Kim, T. Park, i S. Kim, “Batch normalization and bounded activation functions”, u *International Conference on Learning Representations (ICLR)*, 2023. [Mrežno]. Adresa: <https://openreview.net/forum?id=FLr9RRqbwB->
- [31] S. Chen, K. Ma, i Y. Zheng, “Med3d: Transfer learning for 3d medical image analysis”, *arXiv preprint arXiv:1904.00625*, 2019. <https://doi.org/10.48550/arXiv.1904.00625>
- [32] D. Tak, B. A. Gormosa, A. Zapaishchykova *et al.*, “A generalizable foundation model for analysis of human brain mri”, *Nature Neuroscience*, 2026. <https://doi.org/10.1038/s41593-026-02202-6>
- [33] L. E. DeLisi, K. U. Szulc, H. C. Bertisch, M. Majcher, i K. Brown, “Understanding structural brain changes in schizophrenia”, *Dialogues in Clinical Neuroscience*, sv. 8, br. 1, str. 71–78, 2006. <https://doi.org/10.31887/DCNS.2006.8.1/ldelisi>

Izjava o korištenju umjetne inteligencije

Izjava o korištenju umjetne inteligencije: Potvrđujem da sam tijekom izrade ovog diplomskog rada koristila umjetnu inteligenciju u skladu s Politikom primjerenog korištenja umjetne inteligencije na Fakultetu elektrotehnike i računarstva, te da umjetna inteligencija nije zamijenila moje kritičko razmišljanje niti moj doprinos.