

Detekcija tumora mozga sa slika MRI pomoću konvolucijskih neuronskih mreža

Tomislav Krog

Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Zagreb, Hrvatska
tomislav.krog@fer.hr

Rej Šafranko

Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Zagreb, Hrvatska
rej.safranko@fer.hr

Ivan Kušeta

Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Zagreb, Hrvatska
ivan.kuseta@fer.hr

Tihomir Pavić

Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Zagreb, Hrvatska
tihomir.pavic@fer.hr

Tin Ferković

Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Zagreb, Hrvatska
tin.ferkovic@fer.hr

Sažetak: U današnje vrijeme jedno od najvećih problema u medicini su tumori. Tumor mozga vrlo je čest te nažalost u velikoj mjeri završava smrću pacijenta. Iznimno je bitno da se tumor otkrije na vrijeme kako bi se šanse za izlječenje povećale. U te svrhe danas sve veći značaj u medicini doprinosi umjetna inteligencija. U ovom radu opisana su i testirana tri modela dubokog učenja za detekciju tumora mozga: VGG-16 model, ResNet50 model i Vision Transformer (ViT). Svaki model ispitan je na skupu podataka *Br35H:Brain tumor detection 2020*, a rezultati na testnom skupu pokazali su se zadovoljavajućima. Model VGG-16 dao je točnost na testnom setu od 0.963, model ResNet50 točnost od 0.98, dok je ViT dao točnost od 0.995.

Ključne riječi: tumor mozga, duboko učenje, umjetna inteligencija, konvolucijske neuronske mreže

I. UVOD

U današnje vrijeme, tumori su jedna od najopasnijih i najsmrtonosnijih bolesti širom svijeta. Konkretno, tumor mozga smatra se jednim od najopasnijih jer često može dovesti do smrti pacijenta ili ostaviti kobne posljedice. Prema statistici, procjenjuje se da trenutno u SAD-u od tumora mozga boluje 700 000 ljudi, dok svake godine oboli oko 90 000. Također, prema statistici, relativna stopa preživljavanja iznosi 36% [6]. Kako bi se smanjio broj smrtnih slučajeva, potrebno je osnažiti preventivne i dijagnostičke metode te unaprijediti liječenje tumora. Napredak u tehnologiji, poput konvolucijskih neuronskih mreža, ima potencijal za poboljšanje otkrivanja tumora pravovremeno te tako utjecati na uspješnost liječenja i smanjenja štetnih ili smrtnih posljedica.

Konvolucijske neuronske mreže (engl. Convolutional Neural Networks, CNNs) su najčešće korišteni alat za klasifikaciju slika. Sastoje se od više slojeva međusobno povezanih neurona koji procesiraju i transformiraju ulazne podatke kroz učenje. U nastavku ovog rada bit će opisano rješenje za detekciju tumora mozga tj. klasifikaciju MRI slika ovisno o tome postoji li tumor na toj slici ili ne pomoću konvolucijskih neuronskih mreža te modela Vision Transformer koji se ne oslanja na konvolucijsku arhitekturu, međutim, postiže sjajne rezultat. Modeli konvolucijskih neuronskih mreža koji su korišteni u ovom radu su: VGG-16, ResNet50.

II. PREGLED LITERATURE

Postoji mnogo radova koji se bave istim klasifikacijskim problemom koji je opisan u ovom radu. Također, postoje različiti pristupi i modeli dubokog ili strojnog učenja koji mogu riješiti ovaj problem. Konvolucijski modeli su izrazito popularni kod ovakvih problema zbog svoje efikasnosti, a primjer upravo primjene modela VGG-16, koji je korišten i u ovom radu, pokazuje članak [7]. U tom članku autori su dobili točnost na testnom skupu podataka od čak 97.6%. Osim konvolucijskih modela, autori članka [8] su ponudili zanimljivo rješenje napravivši kombinaciju konvolucijskog modela i modela LSTM (unaprijeđen model rekurzivne neuronske mreže). Tim modelom autori su dobili točnost od 99.1%. Osim same klasifikacije, postoji i metoda segmentacije pomoću koje također tumor mozga može detektirati. Primjer segmentacije MRI slika mozga napravili su autori članka [9]. Osim poznati modela, ponekad se na klasifikacijskom problemu poput ovog vrlo dobro ponašaju jednostavniji i manji modeli koji se kreirani od strane autora. Primjer takve mreže može se pronaći ovdje [10], a dobiveni rezultati mogu biti zaista impresivni poput ovog, gdje je dobivena točnost na validacijskom skupu od 99%.

III. OPIS RJEŠENJA

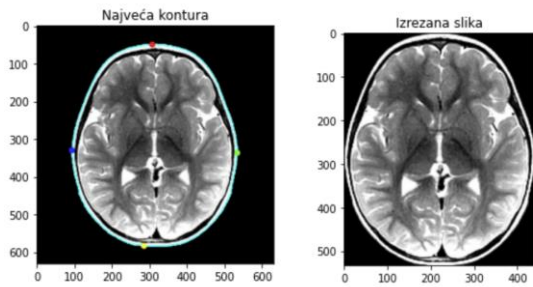
A. Ulazni skup podataka

Ulazni skup podataka *Br35H:Brain tumor detection 2020* [1] sastoji se od 3000 slika mozga dobivenih MRI oslikavanjem. Od ukupnog broja slika, 1500 pripada pacijentima koji imaju tumor mozga, dok preostalih 1500 slika su MRI slike mozga pacijenata na kojima tumor ne postoji.

B. Priprema podataka

Prije samog treniranja neuronske mreže i klasifikacije, potrebno je podatke učitati u odgovarajućem formatu i kako bi dobili što bolje rezultate, nužno je provesti pripremu podataka. Slike su originalno u različitim veličinama te je potrebno promijeniti veličinu tako da je svaka slika iste veličine kako bi se slike mogle predati na ulazni sloj neuronske mreže. Međutim, s obzirom da su originalne slike različitih dimenzije, samom promjenom veličine može doći do izobličenja glavnog objekta sa slike (u ovom slučaju mozga) te bi zbog toga rezultati mogli biti nešto lošiji. Iz tog razloga napravljen je korak predobradbe slike tako da se na slici nastoji detektirati rub mozga te naći 4 specifične točke tog ruba (detektirane konture): najviša, najniža, najdesnija i najljepija točka detektiranog ruba. Nakon pronađenih točaka,

slika se izreže do tih točaka (Slika III-1), a nakon toga promjenom veličine će izobličenja biti puno manja. Slike su promijenjene na veličinu 200x200 piksela.



Slika III-1: Najveća konturu(rub mozga) i izrezana slika

Budući da se radi o binarnoj klasifikaciji, slikama su dodijeljene klase 1 ili 0 ovisno o tome radi li se o slici mozga na kojoj postoji tumor ili ne. Ulazni skup podataka podijeljen je na skupove za treniranje, validaciju i testiranje u omjeru 64:16:20. Dakle, 64% podataka (1920 slika) korišteno je za treniranje mreže tako da se maksimizira točnost na skupu za validaciju koji iznosi 16% podataka (480 slika). Nakon što je model naučen, testiran je na testnom skupu veličine 20% ukupnih podataka (600 slika).

C. Treniranje i klasifikacija

Konvolucijske mreže vrsta su dubokih neuronskih mreža koja postiže posebno dobre rezultate u klasifikaciji slika. Kao i obične neuronske mreže, sastoje se od jednog ulaznog sloja, jednog ili više skrivenih slojeva i izlaznog sloja. Treniranje neuronske mreže je proces prilagođavanja težina neurona na temelju ulaznih podataka i željenih izlaza. U ovom radu klasifikaciju smo isprobali s tri različita modela.

Prvi model je poznati model konvolucijske neuronske mreže VGG-16. Taj model sastoji se od 16 slojeva, a najbitniji su konvolucijski slojevi (koji omogućuju učenje značajki sa slike), slojevi sažimanja (koji služe za smanjenje dimenzije) i potpuno povezani slojevi koji na kraju omogućavaju klasifikaciju [2]. Zbog velike složenosti i velikog broja parametara konvolucijske neuronske mreže te premalog broja slika, odlučili smo koristiti predtrenirani model i trenirati težine samo potpuno povezanog sloja kojeg smo dodali kako bi mreža mogla biti primjenjena na zadani problem, dok će predtrenirani konvolucijski slojevi izdvajati značajke sa slike. Kao aktivacijsku funkciju u potpuno povezanom sloju koristimo ReLu, dok je broj neurona potpuno povezanog sloja 128 (dobiven optimizacijom). Izlazni sloj ima samo jedan neuron koji koristi sigmoidu kao aktivacijsku funkciju. Kao optimizator korišten je optimizator Adam [3] koji je izrazito popularan optimizator u dubokom učenju zbog svoje efikasnosti. Model je treniran na podacima za učenje s različitim veličinama grupa (engl. *batch size*) veličina 16, 32 i 64 te se pokazalo da korištenje grupa od 16 primjera daje najbolju točnost na skupu za validaciju. Uz veličinu grupe, jedan proces učenja imao je 40 epoha, a početna stopa učenja postavljena je na 0.001. Također, okruženje Keras ima mogućnost tzv. *callback* funkcija, a jedna od njih je da se u sklopu učenja smanjuje stopa učenja uz određeni uvjet [4]. U ovom slučaju, ako se točnost na skupu za validaciju nije popravila u 3 uzastopne epohe, stopa učenja se smanjuje na način da se pomnoži faktorom 0.2.

Drugi model koji smo koristili je konvolucijska neuronska mreža ResNet50 predtrenirana na skupu podataka ImageNet. Model je dubok 50 slojeva te svaki sloj ima rezidualnu konekciju. Rezidualna funkcija je aditivna funkcija koja sumira netransformirani ulaz u sloj i izlaz sloja. Time se prenosi nepromijenjena informacija iz prijašnjih slojeva. To je odgovor na problem iščezavajućeg gradijenta kod jako dubokih modela. Arhitektura modela sadrži rezidualne slojeve građene od rezidualnih blokova. Jedan rezidualni blok implementira konvolucijski sloj, sloj sažimanja, sloj normalizacije grupe i ReLu aktivacijski sloj. Na kraju modela je potpuno povezani izlazni sloj koji omogućava klasifikaciju. Izlazni sloj, kao i VGG-16, ima jedan neuron te koristi sigmoidalnu funkciju kako bi modelirao vjerojatnost pozitivne klase. Kao optimizator je također korišten Adam [3] sa stopm učenja 0.001. Koristili smo grupu od 16 primjera za učenje i validaciju. Učenje smo proveli kroz 30 epoha. Postavili smo *EarlyStopping callback* funkciju iz radnog okvira Tensorflow Keras. Parametar *patience* smo postavili na 6 što znači ako se validacijska točnost modela nakon prolaska kroz 6 epoha ne poveća, model se prestaje učiti. To se u našem slučaju dogodilo. Model je postigao najveću validacijsku točnost na 10. epohi te nakon što je i na 16. epohi imao istu validacijsku točnost, prestao se učiti.

Treći model je Vision Transformer (ViT) [11]. Arhitektura Transformer je do tada bila korištena jedino u području obrade prirodnog jezika (NLP) [12]. ViT koristi samo enkoderski dio Transformer, sastoji se od više slojeva, od koji svaki sadrži 3 bloka – normalizacija sloja, mehanizam pozornosti sa više glava te višeslojni perceptron. Ovaj model se, za razliku od prethodnih, ne oslanja na konvolucijske neuronske mreže. Umjesto toga, koristi spomenuti mehanizam pozornosti. Mehanizam pozornosti daje različite težine (važnosti) različitim segmentima slike. Segment slike je dio slike veličine 16x16 piksela, koji je izravnat u vektor prije korištenja. Mehanizam pozornosti ViT-u omogućava da se fokusira na najbitnije dijelove slike s obzirom na zadatak koji rješava. U našem slučaju to su segmenti koji sadrže tumor. ViT nije slučajno inicijaliziran, već su mu težine inicijalizirane predtreniranim modelom, koji je onda dodatno naučen kroz 7 epoha. Veličina grupe za učenje bila je 16, a stopa učenja 0.0002 uz linearno smanjivanje tijekom vremena. Kao optimizator korišten je Adam sa smanjenjem težina [13].

IV. REZULTATI

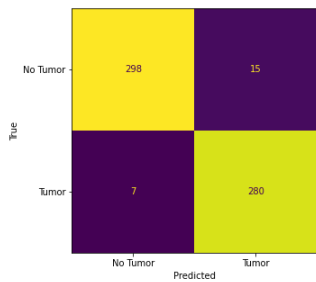
A. Metrike

Za ispitivanje performansi modela koristimo tri metrike: (1) točnost (engl. *accuracy*), (2) odziv modela (engl. *recall*), i (3) F1 ocjenu. Točnost je mjera koja mjeri koliko je ulaznih primjera model ispravno označio. Odziv modela je mjera koja nam govori koliko se pozitivnih primjera odazvalo klasifikatoru [5]. F1 ocjena je harmonijska sredina odziva modela i preciznosti gdje je preciznost omjer broja primjera ispravno označenih pozitivno i ukupnog broja primjera označenih pozitivno [5].

B. Rezultati

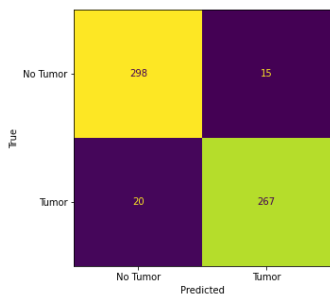
Kod odabira najboljeg modela korištena je mjera točnosti. Model koji se pokazao kao najbolji dao je zadovoljavajuće rezultate (Tablica IV-1). Također, na Slici IV-1 vidimo tzv. konfuzijsku matricu koja nam daje uvid u to kako model klasificira (VGG-16). VGG-16 model je klasificirao 15 slika

u klasu koja predstavlja da tumor postoji, dok te slike zapravo pripadaju klasi koja predstavlja da nema tumora, dok je 7 slika označio kao da tumora na mozgu nema, a on zapravo postoji na slici.



Slika IV-1: Konfuzijska matrica, model VGG-16

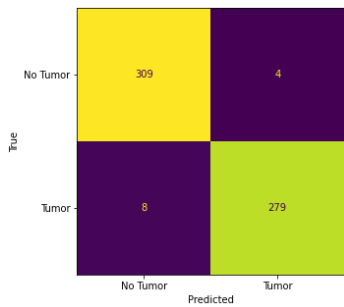
Kako bi pokušali još malo poboljšati rezultate, testiran je isti model, ali na slikama nad kojima je provedena metoda izjednačavanja histograma. Rezultati tog modela također su prikazani u Tablici IV-1, a konfuzijska matrica tog modela prikazana je na Slici IV-2.



Slika IV-2: Konfuzijska matrica, model VGG-16, izjednačavanje histograma

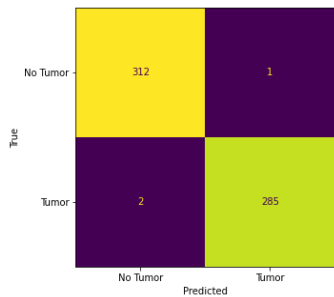
Metoda izjednačavanja histograma nije pomogla u poboljšanju rezultata, štoviše, rezultati su nešto i lošiji.

Rezultati modela ResNet50 su prikazani u Tablici IV-1. Kao najbolji model izabran je onaj koji je dao najbolju točnost na validacijskom skupu podataka te je on ispitan na testnom skupu. Također, na Slici IV-3 vidimo tzv. konfuzijsku matricu koja nam daje uvid u to kako model klasificira. ResNet50 model je klasificirao 4 slike u klasu koja predstavlja da tumor postoji, dok te slike zapravo pripadaju klasi koja predstavlja da nema tumora, dok je 8 slika označio kao da tumora na mozgu nema, a on zapravo postoji na slici.



Slika IV-3: Konfuzijska matrica, model ResNet50

Rezultati ViT-a prikazani su u Tablici IV-1. S obzirom na visoke metrike u ostala dva modela te činjenicu da je rezultate sve teže značajno poboljšavat što su oni bliži maksimumu, prikazani rezultati su izvrsni. U konfuzijskoj matrici na Slici IV-4 vidi se da je model od ukupno 600 slika samo jednu sliku bez tumora klasificirao kao lažno pozitivnu te samo dvije slike s tumorom kao lažno negativne. Iako su to izvrsni rezultati, treba imati na umu da se radi o ljudskim životima te je potrebno izbjeći svaku grešku, posebice lažno negativne predikcije.



Slika IV-4: Konfuzijska matrica, model Vision Transformer

Tablica IV-1: Performanse model

Rezultati	Metrika		
	Točnost	Odziv	F1
VGG-16	0.963	0.976	0.962
VGG-16 izjed-hist	0.941	0.930	0.938
ResNet50	0.980	0.980	0.980
ViT	0.995	0.993	0.995

V. DISKUSIJA

Naši modeli pokazali su slične, ako ne i bolje performanse od drugih radova koji su koristili isti skup podataka. Dobar odabir modela za klasifikaciju slika te iscrpna optimizacija hiperparametara dovela je do toga da svaki model postiže visoki nivo performansi. Manjak neiskustva u oblikovanju ovakvih modela nije nam zadao prevelike probleme te smo uspjeli postići performanse na razini state-of-the-art modela, pogotovo modelom Vision Transformera, premašujući sve unaprijed visoko (ali dostižno) postavljene ciljeve.

VI. ZAKLJUČAK

Detekcija tumora mozga jedan je od ključnih problema u neuroonkologiji. Tumor mozga jedan je od najčešćih i najsmrtonosnijih oblika tumora. Pravovremena dijagnoza i precizna detekcija ključni su faktori za uspješno liječenje. Ovaj članak pokazao je kako su konvolucijske neuronske mreže primjer pouzdane metode za klasifikaciju tumora mozga sa slika te rezultati pokazuju kako ova metoda može značajno pridonijeti ranoj dijagnozi tumora mozga i poboljšati prognozu za pacijente, pogotovo model ResNet50. Ipak, model Vision Transformer koji se pokazao kao najbolji, ne oslanja se na ideju konvolucijske neuronske mreže. Jedna od mogućih implikacija jest automatizacija dijagnoze tumora mozga primjenom konvolucijskih neuronskih mreža ili Vision Transformera, što bi moglo smanjiti terete zdravstvenog sustava i pacijenata.

POPIS LITERATURE

- [1] <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>
- [2] <https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>
- [3] <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
- [4] https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ReduceLROnPlateau
- [5] Šnajder J., *Vrednovanje modela*, Zagreb: Nastavni materijali iz kolegija Strojno učenje 1, Fakultet elektrotehnike i računarstva, 2022.
- [6] <https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/>
- [7] Younis, A.; Qiang, L.; Nyatega, C.O.; Adamu, M.J.; Kawuwa, H.B. Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches. Appl. Sci. 2022, 12, 7282. <https://doi.org/10.3390/app12147282>
- [8] Alsubai S, Khan HU, Alqahtani A, Sha M, Abbas S and Mohammad UG (2022) Ensemble deep learning for brain tumor detection. Front. Comput. Neurosci. 16:1005617. doi: 10.3389/fncom.2022.1005617
- [9] Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghouschi, S. *et al.* Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *SciRep* 11, 10930(2021). <https://doi.org/10.1038/s41598-021-90428-8>
- [10] <https://www.kaggle.com/code/jananiharshu/brain-tumor-detection-simple-cnn>
- [11] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
- [12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
- [13] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.