

Segmentacija tumora jetre

Luka Dikčić
University of Zagreb, FER
Zagreb, Croatia
luka.diktic@fer.hr

Ivan Kovač
University of Zagreb, FER
Zagreb, Croatia
ivan.kovac@fer.hr

Matija Jakovac
University of Zagreb, FER
Zagreb, Croatia
matija.jakovac@fer.hr

Ivan Mihaljević
University of Zagreb, FER
Zagreb, Croatia
ivan.mihaljevic2@fer.hr

Dominik Poljak
University of Zagreb, FER
Zagreb, Croatia
dominik.poljak@fer.hr

Sažetak—Duboko učenje već godinama pomaže pri analizi i obradi slika u biomedicini, a u ovom radu se pomoću njega obrađuje problem segmentacije tumora jetre. Pri tome su korišteni razni modeli temeljeni na konvolucijskim neuronskim mrežama kao što su UNet, SegNet, EfficientNet i transformer. Odabirom njihovih hiperparametara i ispravnom implementacijom tumor se može ispravno segmentirati. Modeli su implementirani pomoću open-source biblioteka TensorFlow i Fastai. Dobiveni rezultati su zadovoljavajući i preklapaju se sa sličnim radovima s obzirom na preciznost i točnost lokacije tumora jetre.

Cljučne riječi—Neuronske mreže, Duboko učenje, CNN, UNet, EfficientNet, SegNet, Transformer, TensorFlow, Fastai

I. UVOD

Tumor jetre 5. je najčešći tumor kod muškaraca te 9. najčešći tumor kod žena te je zbog toga jedan od vodećih problema svjetskog javnog zdravstva. Rana dijagnostika ovih tumora omogućuje raniji početak liječenja što je presudno za poboljšanje kvalitete života i preživljavanje pacijenta. Za razvoj ovakvog modela potrebna je vrlo kvalitetna slika organa odnosno snimka računalne tomografije (CT-a) pripadajućeg abdominalnog područja. Zbog heterogenog i difuznog oblika tumora njegova segmentacija može predstavljati problem ako je uslijed snimanja došlo do smetnji koje nije moguće ukloniti ili ako snimke nisu dovoljno visoke rezolucije.

Upravo zbog složenosti problema, duboko učenje odnosno konvolucijske neuronske mreže (CNN) su idealne za segmentaciju tumora na scan-u zbog njihove mogućnosti učenja raznih karakteristika slika kroz više mogućih slojeva. Uz to, zadržavaju prostornu strukturu slike što je važno radi lokalizacije određenih segmenata odnosno tumora. Neki od modela koji koriste CNN su UNet, SegNet, EfficientNet čiju smo efikasnost ispitali na ovome problemu.

Njihova implementacija se izvodi pomoću open-source biblioteka za duboko učenje TensorFlow koji je razvio Google Brain tim i Fastai koja je razvijena kako bi olakšala rad sa složenim modelima te velikim skupovima podataka. Obje biblioteke sadrže visoko-razvijeni API koji podržava mnoge zadatke uključujući segmentaciju, klasifikaciju, regresiju i sl.

II. KRATAK PREGLED LITERATURE

U pripremi istraživačkog rada kako bi se stvorila čvrsta osnova korišteni su raznoliki izvori. Razne korištene funkcije u izvornom kodu su preuzete sa stranice za dokumentacije biblioteke Fastai [1]. One su nam poslužile pri preprocesuiranju kako bi osigurali da je ulaz u model slika u .jpg formatu. Fastai dokumentacija pruža detaljne smjernice i

kode primjere koji su omogućili učinkovitu pripremu podataka za treniranje modela.

Samo znanje o slikama u medicini odnosno CT scan-u i njegovoj obradi i vizualizaciji pridobili smo proučavanjem članka na online platformi Kaggle [2]. Ovaj članak detaljno opisuje različite tehnike obrade i prikaza DICOM datoteka, što je ključno za razumijevanje strukture i značajki medicinskih slika.

Osim tog znanje, na navedenoj platformi se nalaze već razni postojeći pristupi za rješavanje segmentacije tumora jetre [3] [4] [5] [6]. Ovi pristupi pružaju dragocjene uvide u različite tehnike segmentacije koje su se pokazale uspješnima u praksi. Kaggle nam je također poslužio za skup podataka odnosno za 131 .nii datoteku koja svaka sadrži takozvane slice-ove. Jedan slice predstavlja jedan CT scan područja koje ne mora biti abdominalno [7] [8].

U procesu istraživanja za odgovarajući model pregledali smo razne članke na internetu [9] [10] [11]. Proučavanjem istih odlučili smo za implementaciju određenih modela opisanih u nastavku. Ovi članci pružaju dubinsko razumijevanje različitih arhitektura neuronskih mreža, uključujući UNet i SegNet, te njihove primjene u segmentaciji medicinskih slika.

III. KONVOLUCIJSKE NEURONSKE MREŽE

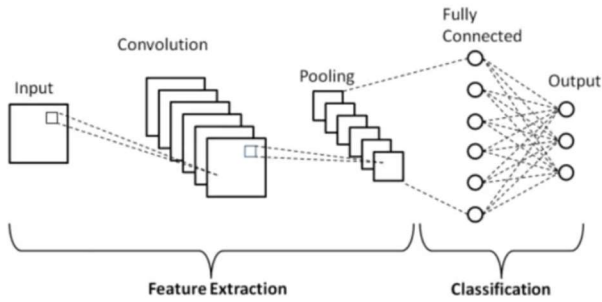
Konvolucijske neuronske mreže posebno su dizajnirane za obradu i analizu vizualnih podataka. Njihov koncept je inspiriran vizualnim sistemom živih organizama. Konvolucijske neuronske mreže (CNN) vrlo su često korišten alat za prepoznavanje slika, u razvoju autonomnih vozila i medicinske dijagnostike.

Struktura CNN-a sastoji se od više slojeva:

- Konvolucijskog sloja; u ovom sloju se primjenjuje operator konvolucije po kojem je neuronske mreže i dobila ime; pomoću filtera otkrivaju se teksture i rubovi.
- Aktivacijskog sloja; u ovom sloju se uvodi nelinearnost i omogućuje mreži da radi složenije poveznice između podataka unutar slike
- Sloja sažimanja (pooling sloj); ovaj sloj se javlja periodično s ciljem da smanji dimenzionalnost podataka i tako pojednostavi računske operacije
- Potpuno povezanog sloja; posljednji sloj u kojem se na temelju obrađenih vrijednosti donosi konačna odluka koja se obično odnosi na klasifikaciju sliku

Općenito, konvolucijske neuronske mreže trenutno su vodeći način implementacije računalnog vida i zbog velike sposobnosti automatskog pronalaženja složenih obrazaca u

podacima imaju ustaljeno mjesto u budućem razvoju umjetne inteligencije.



IV. TRANSFORMER

Transformeri su vrsta arhitekture dubokog učenja koja je prvotno razvijena za obradu prirodnog jezika, ali je ubrzo pronašla primjenu i u drugim granama dubokog učenja. Posebno su pogodni za analizu složenih vizualnih motiva zbog mogućnosti opažanja dugoročnih ovisnosti među elementima unutar podataka.

Sastoje se od dva dijela: enkodera i dekodera; te je u kontekstu računalnog vida važniji enkoder koji se sastoji od sljedećih komponenti:

- Self-attention mehanizam; jedna od ključnih i prepoznatljivih komponenti transformera, služi za izračun pažnje u kontekstu piksela te pomoću te informacije identificira odnose među dijelovima slike i njihovu relevantnost
- Feed-forward mreža; uvodi nelinearnost u sustav i omogućava modelu da spoznaje složene obrasce unutar dobivenih podataka
- Pozicijsko kodiranje; kako transformeri ne koriste konvoluciju, dolazi do potrebe za pozicijskim kodiranjem kako bi se očuvala informacija o prostornom rasporedu piksela

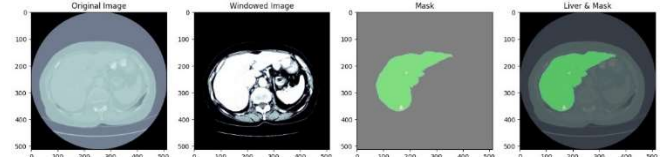
Posebnost transformera je u njihovoj efikasnosti uočavanja globalnih odnosa unutar slike. Također imaju sposobnost prilagodbe i skaliranja sustavno konkretnom problemu za koji se koriste.

V. POČETNI SKUP PODATAKA

Skup podataka korišten u ovom projektu preuzet je s platforme „Kaggle“, a bio je dio natjecanja „LiTS“ (Liver Tumor Segmentation Challenge“ iz 2017. godine. Sve datoteke su u formatu NIfTI (Neuroimaging Informatics Technology Initiative) s ekstenzijom „.nii“. Taj format datoteke se koristi za spremanje medicinskih slikovnih podataka, a za čitanje tog tipa datoteke koristili smo library „nibabel“.

U skupu podataka nalazi se 131 CT snimka abdomena. Naziv svake datoteke odgovara formatu „volume-x.nii“ gdje je x index CT snimke, a naziv odgovarajuće makse odgovara formatu „segmentation-x.nii“ gdje je x indeks CT snimke kojoj maska pripada. Svaka snimka sastoji se od 3 dimenzije, gdje su prve dvije visina i širina slike i one su iste za svaku snimku (iznose 512px), a treća je broj okomitih presjeka abdomena i ovisi od snimke do snimke.

Raspon vrijednosti piksela svakog sloja se razlikuje i puno je veći od klasičnog zato što predstavljaju Hounsfieldovu vrijednost (HU). Pomoću Hounsfieldove skale zapisujemo koliko radijacije određeno tkivo apsorbira i na temelju toga možemo dobiti prikaz unutrašnjosti tijela. Kako ne možemo prikazati sliku zbog velikog raspona vrijednosti piksela, potrebno je primijeniti „windowing“ proces u kojem sliku mijenjamo na temelju određenog prozora odnosno na temelju parametara razine i širine. Vrijednosti prozora za određene strukture su unaprijed poznate (za jetru je širina 150, a razina 30). Na kraju dobivamo prikaz slike i maske.



Kako bi skup podataka mogli primijeniti za treniranje modela dubokog učenja, razdvajamo CT snimku na razine i spremamo u novi skup podataka. Algoritam kreiranja novog skupa podataka dan je u Algoritmu 1.

Algoritam 1: Izgradnja novog skupa podataka

```

Ulaz: df - Skup putanja do CT slike i njihovih maski
for file_info in df do
    slika_3D := učitaj_3D_sliku(file_info)
    maksa_3D := učitaj_3D_masku(file_info)
    slojevi := broj_slojeva(slika)
    for sloj_index from 0 to slojevi do
        slika := PRETVORI odaberi_sloj(slika[sloj_index]) U FLOAT32 tensor
        maska := PRETVORI odaberi_sloj(maksa[sloj_index]) U FLOAT32
        windowed := window(slika, širina, razina) // windowing proces
        bins := izjednači_histogram(slika) // Računa granice histograma
        skaliran := skaliraj_tensor(slika, bins) // skaliranje na temelju granica
        slika := stack(windowed, skaliran) // prozorirana i skalirana zajedno
        spremi(slika)
        spremi(maska)
    end
end

```

Od početnog skupa podataka koji je sadržavao 131 datoteku, stvorili smo novi koji u 'train' skupu sadrži 17891 sliku, u 'validation' skupu 5309, a u 'test' skupu 6156

VI. METODOLOGIJA

A. UNet

Jedan od modela koji se koristio prilikom segmentacije je UNet koji je jedan od najčešće korišten modela u biomedicini. Svoje ime je dobio po samom izgledu strukture koje nalikuje na slovo U. Sastoji se od dva glavna dijela, enkodera te dekodera uz pomoć skip-konekcijama i bottleneck sloja.

Enkoder se sastoji od više konvolucijskih blokova koji su izgledom isti te se sastoje od 4 komponente. Kao u većini modela, ulaz je slika koja zatim prolazi kroz jednu dvodimenzionalnu konvoluciju, zatim se normalizira te se na kraju primjenjuje aktivacijska funkcija ReLU. Taj postupak se ponovi još jedanput te se blok završava primjenjivanjem

max-pooling funkcije koja smanjuje dimenziju slike pritom zadržavajući ključne karakteristike. Takva slika se predaje sljedećem bloku u nizu kojih ukupno ima 4. Ključni enkoderv zadatak je izdvojiti bitne karakteristike slike iz ulazne slike pritom smanjujući njezinu dimenziju.

Conv2D -> Normalization -> ReLU -> Conv2D ->
Normalization -> ReLU -> MaxPooling

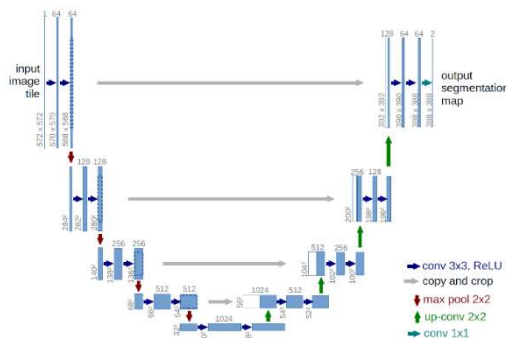
slok je dio modela koji sadrži najviše i najkompleksnije karakteristike slike unatoč najmanjoj dimenziji slike u modelu. On je također zaslužan za spoj enkodera i dekodera te se također sastoji od dvije konvolucije s normalizacijom i istom aktivacijskom funkcijom.

Conv2D -> Normalization -> ReLU -> Conv2D ->
Normalization -> ReLU

Dekoder započinje povećanjem rezolucije slike metodom koja se naziva „up-sampling“. Nakon povećanje rezolucije koriste se skip-konekcije koje spajaju izlaz enkoderskog s ulazom dekoderskog bloka tako omogućavajući točno lociranje specifične karakteristike slike. Ponovno se izvršava dvije već spomenute konvolucije s aktivacijom.

Upsampling -> Concatenate -> Conv2D -> Normalization ->
ReLU -> Conv2D -> Normalization -> ReLU

Izlazni sloj koristi „softmax“ aktivaciju za stvaranje slike koja je segmentirana u tri klase – pozadina, jetra i tumor.



B. SegNet

Model koji se također koristi u semantičkoj segmentaciji je SegNet. Njegova memorijska efikasnost i jednostavnost implementacije ga čine poznatim u industriji. Što se tiče njegove arhitekture ne razlikuje se previše od UNeta osim na način spajanja enkoderskih i dekoderskih blokova. Ne koriste kompleksne skip-konekcije već se oslanja na indeksiranje odnosno zadržavanje indeksa pozicija maksimalnih vrijednosti tijekom primjere max-pooling funkcije.

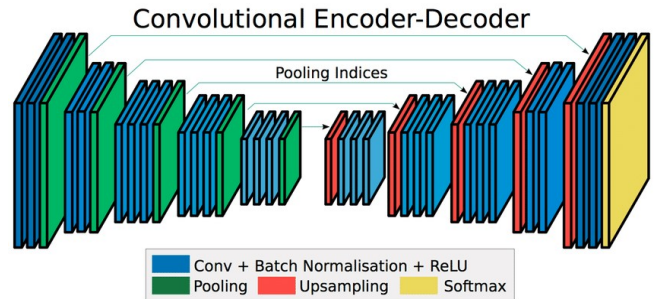
Kao i u UNetu primjenjuje se konvolucija s normalizacijom i ReLU aktivacijskom funkcijom. Naposljetku se primjenjuje već spomenuta max-pooling funkcija. Enkoder se sastoji od 4 bloka.

Conv2D -> Normalization -> ReLU -> MaxPooling

Kako nije bilo promjena kod enkodera, tako nema ni kod dekodera. On se također sastoji od 4 bloka primjenjujući „up-sampling“, konvoluciju, normalizaciju i aktivaciju.

Upsampling -> Conv2D -> Normalization -> ReLU

Izlazna aktivacija je također segmentacije koja ponovno generira sliku segmentiranu na 3 klase.



U oba modela (UNet i SegNet) su se pri treningu također dvije evaluacijske metrike, DiceCoefficient i Precision.

DiceCoefficient se kao evaluacijska metrika koristi pri zadatku segmentaciju te uspoređuje predikcije i stvarnu masku. Specifičnije, računa se kao dvostruka veličina presjeka predviđene i stvarne segmentacije podijeljeno s brojem piksela u obje segmentacije (1). Interval njegovih vrijednosti je od 0 do 1, gdje 1 predstavlja savršeno preklapanje te 0 nikakvo preklapanje.

Precision je jednostavna metrika koja se računa kao razlomak broja ispravno predviđenih primjera i sume istih i neispravno predviđenih primjera. Interval vrijednosti je također od 0 do 1 gdje 1 predstavlja savršenu segmentaciju.

C. EfficientNet

Treći model koji smo koristili za semantičku segmentaciju tumora jetre je EfficientNet. To je konvolucijska neuronska mreža koja koristi metode složenog skaliranja.

Glavna ideja je da se skaliraju 3 glavne dimenzije mreže: dubina, širina i rezolucija jer time postizemo optimalnost kod veličine modela, njegove točnosti i složenosti. EfficientNet je obitelj 8 modela koji su nazvani od B0 do B7 gdje se svaki razlikuje po svojim glavnim parametrima.

U ovom projektu koristili smo najmanji od njih (B0) treniran na skupu „ImageNet“ (skup slika od 1000 razreda). Unaprijed trenirane težine smo zamrznuli kako ih ne bi poremetili tijekom treniranja i koristili kao bazni dio modela. Taj bazni dio modela nam samo služi kao ekstraktor značajki i na njega dodajemo dodatne slojeve koji će iskoristiti prepoznate značajke.

Dodajemo slojeve za uzorkovanje koji vraćaju prostornu rezoluciju značajki na originalnu veličinu slike, konvoluciju i na kraju sloj s aktivacijskom funkcijom softmax za predviđanje razreda za svaki piksel u ulaznoj slici.

D. Transformer

Četvrti model na kojim smo evaluirali točnost semantičke segmentacije tumora jetre u svojoj arhitekturi koristi transformere.

Arhitektura ovoga modela sastoji se od nekoliko dijelova:

- Normalizacija na srednju vrijednost 0 i varijancu 1

- Pretvaranje tenzora slike u skup dijelova dimenzija 16x16
- Kreiranje vektora značajki za svaki dio slike i kreiranje vektora značajki ovisno o poziciji. Ove dva vektora se zbroje kako bi bili ulaz u sloj enkodera transformatora.
- Transformator enkoder
- Linearan sloj koji služi za predikciju
- Sloj koji pretvara sliku iz njezinih dijelova nazad u cjelinu

VII. REZULTATI

A. UNet

Za treniranje UNet modela koristila su se dva skupa podataka, za treniranje te validaciju. Skup za treniranje je sadržavao 8000 nasumično odabranih slika od ukupno 17891 iz originalnog skupa radi smanjenja trajanja treninga. Skup za validaciju je sadržavao 5309 slika. Hiperparametri modela su bili sljedeći:

- Stopa učenja: 0.001
- Veličina serije (batch size): 16
- Broj epoha: 3

Zadnji hiperparametar koji je korišten je dodjeljivanje različitih vrijednosti težina za klase kako bi se ujednačila zastupljenost pojedine klase. Vrijednosti težine su bile sljedeće:

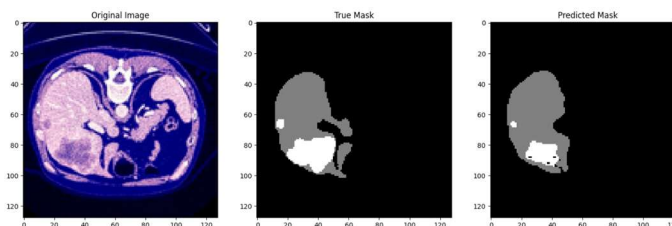
- Pozadina (klasa 0): 1
- Jetra (klasa 1): 5
- Tumor (klasa 2): 20

Koristio se optimizator Adam (Adaptive Moment Estimation) koji služi za prilagođavanje težina u modelu te koristi kombinaciju RMSProp (Root Square Mean Propagation) i SGD (Stochastic Gradient Descent). Funkcija gubitka je izračunata pomoću kategorične unakrsne entropije (eng. Categorical crossentropy)

Nakon tri epohe treninga gdje se pritom svake epohe model poboljšavao, primijenjen na testni skup podataka od 6159 slika. Zabilježeni su sljedeći rezultati:

- Dice score: 0.9846
- Precision: 0.9953

Samom analizom brojki rezultati su i više nego zadovoljavajući te to potvrđuje i vizualna prezentacija na jednoj slici testnog skupa.

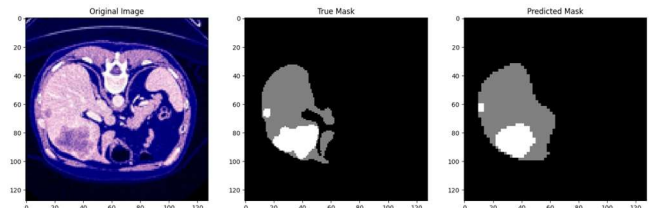


B. SegNet

Kako nije bilo prevelike razlike u implementaciji modela između UNet-a i SegNet-a tako nema ni prevelike razlike u treniranju modela. Sva tri skupa podataka ostaju isti, kao i optimizator te funkcija gubitka. Jedina razlika je hiperparametar broja epoha koji iznosi 5, međutim ostali hiperparametri su ostali isti kao i evaluacijske metrike. Nakon 5 epoha treninga zabilježeni su sljedeći rezultati na testnom skupu podataka.

- Dice score: 0.9899
- Precision: 0.9908

Rezultati su također i više nego zadovoljavajući što dokazuje i vizualizacija na istoj slici testnog skupa podataka kao i kod UNet-a.



C. EfficientNet

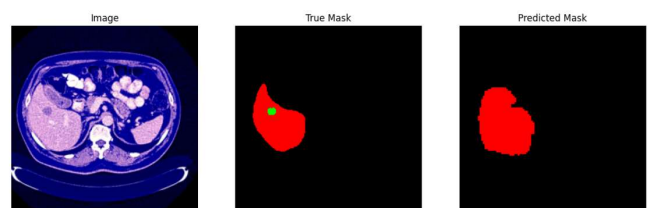
Za treniranje modela kojem je bazni dio bio EfficientNetB0, koristili smo cijeli skup podataka. Hiperparametri modela su bili sljedeći:

- Stopa učenja: 0.001
- Veličina serije: 64
- Broj epoha: 20

Funkcija gubitka je 'categorical_crossentropy', a optimizacijski algoritam 'Adam'. Tijekom treniranja, korištene su accuracy i dice_score kao metrike. Nakon treninga, najbolji rezultati na validation skupu su:

- Dice score: 0.944
- Accuracy: 0.9502
- Validation_loss: 0.1351

Kako smo koristili najmanji EfficientNet model čije smo težine zamrznuili, broj treniranih težina je oko 7 milijuna. Iako su iznosi metrika vrlo dobri, rezultati koje model predviđa na test skupu nisu zadovoljavajući.



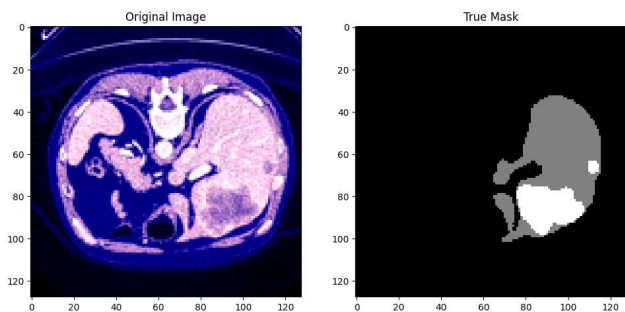
D. Transformer

Za treniranje transformera korišten je već implementirani model [12] na 8000 nasumično odabranih slika radi bržeg učenja. Hiperparametri modela bili su sljedeći:

- Stopa učenja: 0.0004
- Veličina serije (batch size): 64
- Broj epoha: 10

Kao funkcija gubitka korištena je kategorička unakrsna entropija, a kao optimizator Adam. Prilikom izračunavanja metrika na skupu za testiranje dobiveni su sljedeći rezultati:

- Dice score: 0.9921
- Accuracy: 0.9948



VIII. USPOREDBA MODELA

Iz prethodno dobivenih rezultata možemo vidjeti da su UNet i SegNet postigli vrlo dobre rezultate, kao i transformer, međutim EfficientNet nije dobro generalizirao na testnom skupu te zbog toga ne daje dobre segmentacije. Razmatrajući dice score kao metriku po kojoj bi se modeli mogli usporediti, transformer postiže najbolje rezultate, no UNet i SegNet nisu puno lošiji te bi možda uz veći broj epoha učenja i drugačiju stopu učenja mogli biti bolji od transformera.

Kada bismo naše modele uspoređivali s modelima drugih javnih radova vidimo da postižu jednake rezultate kao i oni. Rezultati vanjskih radova također se kreću oko 98 – 99 posto točnosti. [13]

IX. ZAKLJUČAK

U ovom radu istražili smo primjenu dubokog učenja za segmentaciju tumora jetre koristeći različite modele temeljene na konvolucijskim neuronskim mrežama i transformerima. Uspješno smo implementirali i testirali četiri modela: UNet, SegNet, EfficientNet i model baziran na transformerima,

Svi gore navedeni modeli postigli su visoku točnost na skupu za testiranje što ukazuje na to da je ovaj problem pogodan za rješavanje dubokim modelima. Međutim, iako bi ovolika točnost možda bila zadovoljavajuća u nekim drugim vrstama primjene, ona je još uvijek preniska kako bi se koristila u kliničke svrhe. Daljnji razvoj samih algoritama dubokog učenja zasigurno će dovesti i do točnijih modela iz područja segmentacije u medicini koji će se moći koristiti u praksi što će uvelike doprinjeti prevenciji raka jetre i pomoći liječnicima pri otkrivanju ove zloćudne bolesti.

LITERATURA

- [1] J. Howard et al., "fastai Medical Imaging Documentation," available: <https://docs.fast.ai/medical.imaging.html>
- [2] R. Sharma, "CT Scans DICOM Files Windowing Explained," available: <https://www.kaggle.com/code/redwankarimsony/ct-scans-dicom-files-windowing-explained>
- [3] R. Spark, "Liver Tumor Segmentation Final," available: <https://www.kaggle.com/code/rishabhspark/liver-tumor-segmentation-final>
- [4] S. Tenebris, "Liver Segmentation ResNet-50," available: <https://www.kaggle.com/code/tenebris97/liver-segmentation-resnet-50>
- [5] L. Luyen, "Liver Segmentation UNet," available: <https://www.kaggle.com/code/luyen0/liver-segmentation-unet>
- [6] N. Sowmik, "Liver Segmentation with fastai v2 using UNet," available: <https://www.kaggle.com/code/nurislamsowmik/liver-seg-with-fastai-v2-using-unet-sowmik>
- [7] A. Mvd, "Liver Tumor Segmentation Dataset," available: <https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation/data>
- [8] A. Mvd, "Liver Tumor Segmentation Part 2 Dataset," available: <https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation-part-2>
- [9] S. Haider, "SegNet," available: <https://medium.com/@saba99/segnet-a139ce77b570#:~:text=1,identifying%20objects%20and%20their%20boundaries>
- [10] A. Vidhya, "What is UNet," available: <https://medium.com/analytics-vidhya/what-is-unet-157314c87634>
- [11] A. Sarkar, "Understanding EfficientNet: The Most Powerful CNN Architecture," available: <https://arjun-sarkar786.medium.com/understanding-efficientnet-the-most-powerful-cnn-architecture-eaeb40386fad>
- [12] D. Matani, "Efficient Image Segmentation Using PyTorch: Part 4", <https://towardsdatascience.com/efficient-image-segmentation-using-pytorch-part-4-6c86da083432>
- [13] Rahman H, Bukht TFN, Imran A, Tariq J, Tu S, Alzahrani A. A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. Bioengineering (Basel). 2022 Aug