

Klasifikacija melanoma na slikama kožnih lezija

Zrinka Pećanić
Fakultet elektrotehnike i računarstva
Zagreb, Hrvatska
zrinka.pecanic@fer.hr

Marin Avirović
Fakultet elektrotehnike i računarstva
Zagreb, Hrvatska
marin.avirovic@fer.hr

Tomislav Prhat
Fakultet elektrotehnike i računarstva
Zagreb, Hrvatska
tomislav.prhat@fer.hr

Ivana Krišto
Fakultet elektrotehnike i računarstva
Zagreb, Hrvatska
ivana.kristo@fer.hr

Dominik Pavel
Fakultet elektrotehnike i računarstva
Zagreb, Hrvatska
dominik.pavel@fer.hr

Sažetak—Ovaj rad obrađuje problematiku raka kože, s fokusom na melanom, koji je najsmrtonosniji oblik bolesti. Napominje da je rano otkrivanje ključ uspješnog liječenja i sugerira da bi korištenje analize slike temeljene na umjetnoj inteligenciji moglo biti koristan alat u ovom procesu. U radu se posebno istražuje korištenje modela strojnog učenja za klasifikaciju melanoma prema fotografijama kožnih lezija.

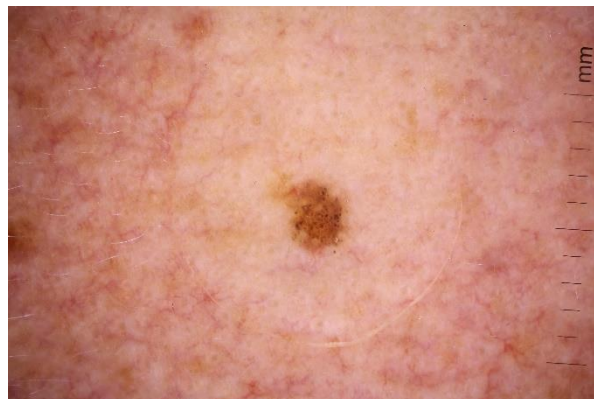
Ključne riječi— rak kože, melanom, smrtonosna bolest, rano otkrivanje, kirurgija niskog rizika, dermatolozi, pregled, kožne lezije, madeži, umjetna inteligencija, analiza slike, modeli strojnog učenja, fotografije, klasifikacija

I. UVOD

Rak kože najčešći je tip raka. Specifični tip istoga je melanom koji je uzrok čak 75% smrtnih slučajeva kod osoba oboljelih od raka kože. Iako je melanom po život opasna bolest, ako se otkrije na vrijeme, melanom se može ukloniti niskorizičnim kirurškim zahvatom. Pravovremeno otkrivanje obavljaju dermatolozi detaljnim pregledom kožnih lezija i madeža, i upravo u tom koraku bismo mogli unaprijediti dijagnostiku ove bolesti koristeći principe umjetne inteligencije. Analizom fotografija kožnih lezija koristeći algoritme za detekciju melanoma može se smanjiti teret i ubrzati posao zdravstvenih djelatnika. Ovaj rad fokusira se na klasifikaciju melanoma iz fotografija kožnih lezija koristeći analizu slike uz primjene modela strojnog učenja.



Slika 1.1 Maligna kožna lezija



Slika 1.2 Benigna kožna lezija

II. PREGLED POSTOJEĆIH RJEŠENJA

Osnova našeg rješenja ovog problema bilo je natjecanje za izradu najboljeg modela za klasifikaciju melanoma organizacije SIIM (Society for Imaging Informatics in Medicine) u suradnji s ISIC (International Skin Imaging Collaboration) čija arhiva sadrži najveću javno dostupnu bazu podataka s fotografijama kožnih lezija. Slike su dostupne u DICOM formatu, što je često korišteni format medicinskih slikovnih podataka. Slike su također dostupne u JPEG i TFRecord formatu. Veličina slika u formatu TFRecord promijenjena je na jedinstvenu 1024x1024. Metapodaci se također pružaju izvan DICOM formata, u CSV datotekama. Podaci za treniranje imaju 8 značajki i 33126 primjera, a podaci za testiranje 5 značajki i 10982 primjera.

Značajke u skupu za treniranje su ime slike, identifikacijski broj pacijenta, spol, aproksimacija dobi, dijagnoza, zastavica za benigni ili maligni melanom i cilj.

S druge strane, značajke u skupu za testiranje su ime slike, identifikacijski broj pacijenta, spol i aproksimacija dobi.

Zadatak natjecanja bio je napraviti i istrenirati model na tim slikama te ga primijeniti na grupu neoznačenih slika. Model treba svakoj slici dodijeliti vrijednost između 0 i 1, gdje 0 označava benignu kožnu leziju, a 1 označava melanom. Predikcije za neoznačene slike (uz ime slike) bilo je potrebno spremati u tablicu submission.csv te se ona (uz kompletnu bilježnicu koja sadrži sav potreban kôd) vrednuje kao konačan rezultat u natjecanju.

Među najboljima u natjecanju odabrali smo pet modela te ih analizirali i/ili modificirali kako bismo usporedili alternativne rezultate i vidjeli kako određene promjene utječu na njihove performanse.

Odabrani modeli jesu sljedeći:

Model M1:

<https://www.kaggle.com/code/amyjang/tensorflow-transfer-learning-melanoma>

Model M2:

<https://www.kaggle.com/code/haqishen/1st-place-soluiton-code-small-ver/notebook>

Model M3:

<https://www.kaggle.com/code/abhishek/melanoma-detection-with-pytorch>

Model M4:

<https://www.kaggle.com/code/saife245/melanoma-detail-analysis-eda-ip-augmentation-model/notebook#Model>

Model M5:

<https://www.kaggle.com/code/datafan07/analysis-of-melanoma-metadata-and-effnet-ensemble/notebook>

III. OPIS RJEŠENJA PROJEKTOG TIMA

A. Model M1

Model korišten za predikciju je VGG16 model. To je već istreniran model te mu ne mijenjamo težine, ali ne uključujemo slojeve s vrha (jer ne radimo klasifikaciju na istim slikama na kojima je model treniran).

Postavljamo početni bias kao $\log(\text{br. malignih/br. benignih})$ jer želimo da naš izlaz reflektira disbalans koji imamo u podacima - jako puno benignih primjera, vrlo malo malignih primjera.

Iz istog razloga želimo da maligne slike imaju veću težinu u modelu, pa i to postavljamo.

Slike su dekodirane i normalizirane tako da je vrijednost svakog piksela između 0 i 1 (umjesto 0-255).

Augmentacija prije treniranja je dobivena nasumičnom rotacijom slika ulijevo. Originalni model treniran je s augmentiranim podacima, a drugi (modificirani) je treniran bez augmentacije.

Ne koristimo accuracy (točnost) kao mjeritelj kvalitete modela (isti razlog kao i prije - disbalans malignih i benignih primjera), već površinu ispod ROC krivulje (`keras.metrics.AUC`).

B. Model M2

Model koji se koristi za predikciju je `enetr2`. Za treniranje radimo tzv. 5-fold training. Svaki nabor (fold) ima 5 epoha (ukupno 25). Tako istrenirani model koristi se dalje za predikciju.

Augmentacija prije treniranja je dobivena rotiranjem slika te promjenom njihove svjetline i kontrasta.

Originalni model je treniran s augmentiranim slikama, a drugi (modificirani) je treniran bez augmentacije.

C. Model M3

SE-ResNeXt-50 je arhitektura korištena u modelu M3. To je arhitektura duboke neuronske mreže koja se temelji na ResNeXt arhitekturi i koristi tehniku nazvanu Squeeze-and-

Excitation (SE) za poboljšanje performansi. "50" u nazivu odnosi se na broj slojeva u modelu. SE-ResNeXt-50 je moćna arhitektura koja se obično koristi za zadatke klasifikacije slika.

Augmentacija u originalnom modelu dobivena je nasumičnim translacijom, skaliranjem, rotiranjem i zrcaljenjem slika. Modificirani model treniran je bez augmentacije.

D. Model M4

Model koji se koristi za predikciju je sekvencijski Xception. Ovaj model je gotov model i predviđa vjerojatnost između 0 i 1 da je lezija na slici maligna. Primijenjena je obrada slike i augmentacija. Korišteno je smanjenje buke, Gaussovo zamućenje, prilagodba kontrasta, adaptivni prag, binarno određivanje praga, izračunavanje površine i parametra kancerogenog dijela stanice, povećanje i smanjivanje piksela slika, ORB (orijentiran FAST i rotirani BRIEF). Treniranje se odvija u 10 epoha.

Budući da je ovo jedan od najbolje ocijenjenih modela na natjecanju, odlučili smo ga obraditi iako mu zbog kompleksnosti izvedbe nije bilo moguće ukloniti augmentaciju.

Za tehnike augmentacije korišteno je navedeno: Cutout, Cutmix i Gridmask data augmentacija.

E. Model M5

Korišteni model je EffNet model. Model prvo skalira sve slike na istu veličinu, zatim ga pustimo da trenira 5-fold stratifikaciju za CV (Cross-Validation). Nakon CV-a model izbacuje meta rezultate (ovo se ne mijenja s augmentacijom).

Zatim model radi adversarijalnu validaciju u kojoj se ciljevi za oba skupa podataka postavljaju na 0 za trening i na 1 za testiranje. Time dobivamo zadnje tablične podatke, koji nisu vezani uz augmentaciju modela.

Augmentacija se sastoji od rotacija, smicanja, povećanje visine i širine te pomaka visine i širine te se odabire nasumično za svaki od primjera. Trening se odvija u 12 epoha nakon kojih dobivamo konačne rezultate. Modificirani model treniran je bez augmentacije.

IV. EKSPERIMENTALNI REZULTATI

Podatci s kojima smo trenirali modele nisu dobro balansirani – od nešto više od trideset i tri tisuće slika, njih samo 584 sadrži melanom. Taj problem riješen je augmentacijom tj. umjetnim nadopunjavanjem dobivenim transformacijom slika iz početnih podataka. Svi originalni modeli sadrže neki oblik augmentacije podataka, a 4 od 5 modela smo modificirali tako da se mogu trenirati i bez augmentacije.

Rezultate uspoređujemo po prikladnim kriterijima – izgledu podataka u tablici `submission.csv` koju, zbog velikog broja podataka u tablici, vizualiziramo histogramom i ROC_AUC vrijednosti koju smo koristili za vrednovanje kvalitete modela. Točnost (accuracy) ovdje nije prikladno koristiti zbog disbalansa u skupu podataka za treniranje, a i zato što nam ne odgovara prirodi problema. Primjerice, ako imamo 100 pacijenata i njih 99 ima benignu kožnu leziju, a jedan malignu, i naš model daje kao rezultat da svi pacijenti imaju benignu kožnu leziju – tada imamo točnost predikcije

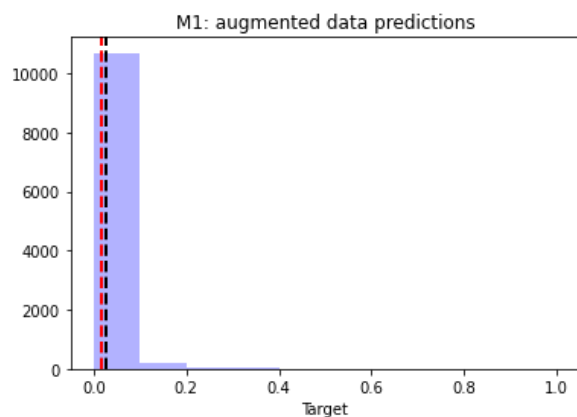
99% (jako visoko i ‘dobro’) no jedan pacijent će imati neotkriveni melanom što ga može koštati zdravlja, ali i života. Stoga koristimo ROC i AUC kao mjeritelje kvalitete predikcija. ROC krivulja (Receiver Operating Characteristic curve) je graf koji prikazuje 2 parametra: učestalost točno klasificiranih primjera i učestalost lažno pozitivnih primjera. AUC (Area under the ROC Curve) je površina ispod te krivulje i u rasponu je od 0 do 1, s time da vrijednosti bliže 0 znače lošu, a bliže 1 dobru klasifikaciju. Ovakav ‘manje crno-bijeli’ pristup nam puno bolje pristaje uz prirodu problema kojim se bavimo.

Slijede odabrani prikazi rezultata za svaki od modela. Crvena linija na histogramima predstavlja medijan, a crna srednju vrijednost.

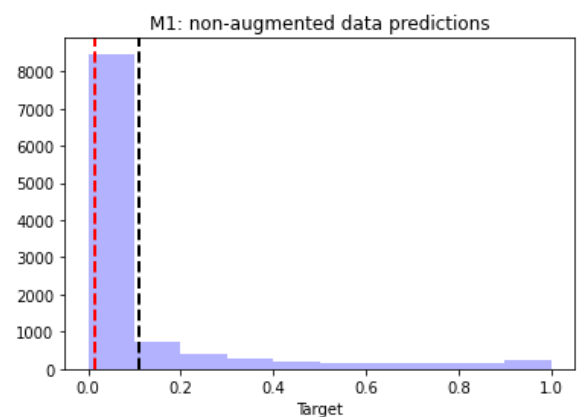
Model M1

AUC nakon treniranja s augmentiranim skupom podataka: 0.8420

AUC nakon treniranja s ne-augmentiranim skupom podataka: 0.8143



Slika IV.1 Predikcije M1 uz augmentaciju



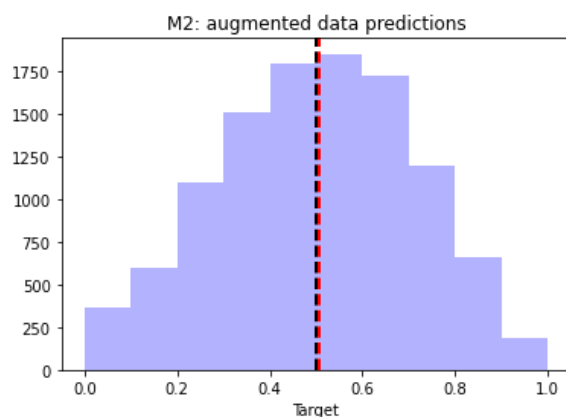
Slika IV.2 Predikcije M1 bez augmentacije

Mode

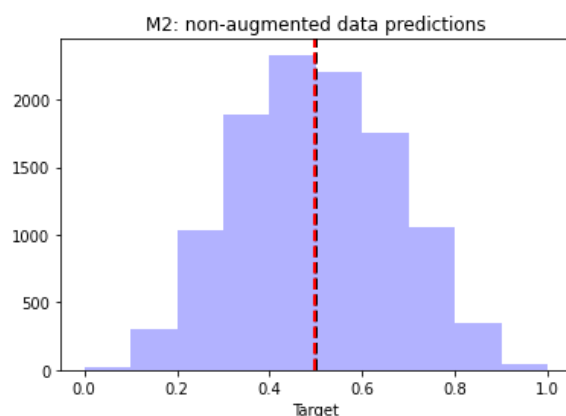
Model M2

AUC nakon treniranja s augmentiranim skupom podataka: 0.7613

AUC nakon treniranja s ne-augmentiranim skupom podataka: 0.7128



Slika IV.3 Predikcije M2 uz augmentaciju



Slika IV.4 Predikcije M2 bez augmentacije

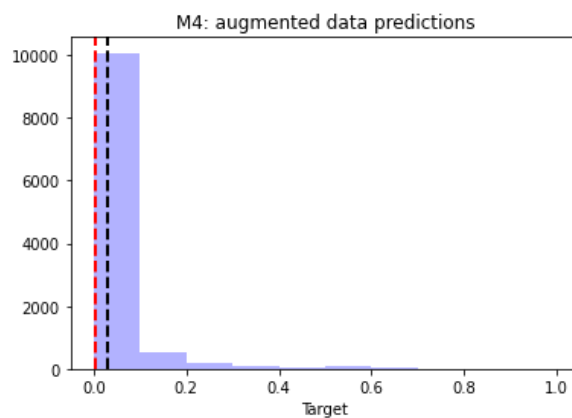
Model M3

AUC nakon treniranja s augmentiranim skupom podataka: 0.8937

AUC nakon treniranja s ne-augmentiranim skupom podataka: 0.8529

Model M4

AUC nakon treniranja s augmentiranim skupom podataka: 0.9437

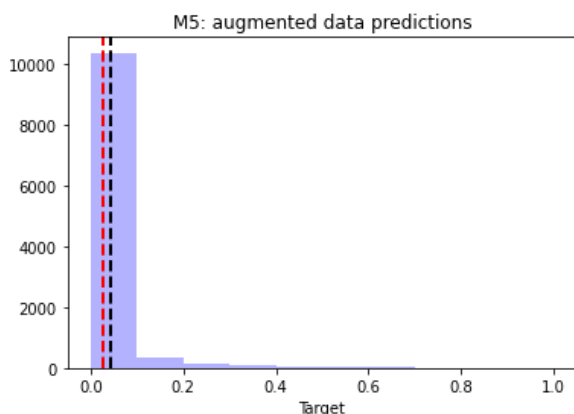


Slika IV.5 Predikcije M4 uz augmentaciju

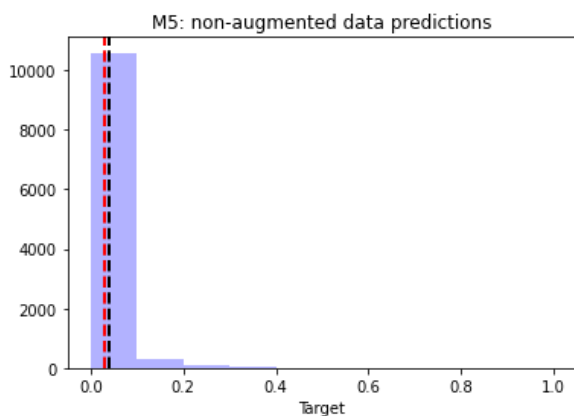
Model M5

AUC nakon treniranja s augmentiranim skupom podataka: 0.9997

AUC nakon treniranja s ne-augmentiranim skupom podataka: 0.996



Slika IV.6 Predikcije M5 uz augmentaciju



Slika IV.7 Predikcije M5 bez augmentacije

V. DISKUSIJA REZULTATA

Očekujemo bolje rezultate koristeći augmentaciju, zbog prethodno navedenog disbalansa među slikama za treniranje. Također, ne znamo točne predikcije za skup za testiranje, no možemo pretpostaviti (s obzirom na distribuciju malignih i benignih primjera u skupu za treniranje) da su velika većina (ako ne i svi) testni primjeri benigne kožne lezije.

S obzirom na dobivene AUC vrijednosti i histograme, vidljivo je da najbolju predikciju rade modeli M4 i M5. Najlošiji model je M2, čija je distribucija predikcija otprilike centrirana oko 0.5 što je nezadovoljavajuće budući da želimo rezultate blizu 0 ili 1 (što manje nejasnih interpretacija koje su oko 0.5). Modeli M1 i M3 se imaju solidne performanse, no nisu toliko dobri M4 i M5. M3 kao predikciju daje negativne vrijednosti umjesto vrijednosti između 0 i 1, no mi smo ga svejedno koristili budući da je priznat na natjecanju (i ima visoke ocjene), a ima i prihvatljive AUC vrijednosti.

Svi modeli rade kvalitetniju predikciju (veća AUC vrijednost) na augmentiranom skupu podataka, što smo i očekivali.

VI. ZAKLJUČAK

Motivirani problematikom ranog otkrivanja melanoma proučili smo natjecanje organizacije SIIM (Society for Imaging Informatics in Medicine) u suradnji s ISIC (International Skin Imaging Collaboration) i odabrali 5 rješenja među onim najboljima. Njih smo zatim analizirali i/ili modificirali kako bismo produbili vlastito znanje o analizi slika u medicini i njezinoj primjeni u dijagnostici. Radili smo s vrlo disbalansiranim podacima za treniranje zbog čega je bilo važno augmentirati podatke za treniranje i koristiti AUC vrijednost umjesto poznatije točnosti (accuracy) za vrednovanje modela. Od odabranih 5 modela koji svi implementiraju različite modele i vrste augmentacija, najbolje performanse su imali M4 i M5, zatim M1 i M3, a najlošije rezultate dobili smo s modelom M2. Daljnji rad u ovom smjeru uključivao bi poboljšanje modela i njegovo eventualno testiranje na nekim novijim slikama kožnih lezija koje nisu dio danog nam skupa podataka.

LITERATURA

- [1] <https://www.kaggle.com/competitions/siim-isic-melanoma-classification/code?competitionId=20270&sortBy=voteCount>, natjecanje korišteno za izbor modela i podatke
- [2] <https://www.kaggle.com/code/amyjang/tensorflow-transfer-learning-melanoma>, model M1
- [3] <https://www.kaggle.com/code/haqishen/1st-place-solution-code-small-ver/notebook>, model M2
- [4] <https://www.kaggle.com/code/abhishek/melanoma-detection-with-pytorch>, model M3
- [5] <https://www.kaggle.com/code/saife245/melanoma-detail-analysis-eda-ip-augmentation-model/notebook#Model>, model M4
- [6] <https://www.kaggle.com/code/datafan07/analysis-of-melanoma-metadata-and-effnet-ensemble/notebook>, model M5