

**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU**

**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMATIKI  
TEHNOLOGIJA OSIJEK**

**Sveučilišni studij računarstva**

**Klasifikacija medicinskih slika za dijagnostiku kožnih  
bolesti**

**Projektni zadatak**

**Patricia Adžić, Samuel Adžić**

**Osijek, 2025.**

# SADRŽAJ

<b>1. UVOD</b>	<b>1</b>
<b>2. SKUP PODATAKA I ANALIZA</b>	<b>2</b>
2.1 Opis skupa podataka	2
2.1.1 Klase	2
2.1.2 Oznake i primjeri podataka	2
2.2 Podjela na skupove za trening i testiranje	3
2.3 Rezolucija i format slika	3
2.4 Eksplorativna analiza podataka	3
2.5 Predobrada i augmentacija podataka	4
<b>3. OPIS KORIŠTENIH METODA I ALGORITAMA</b>	<b>7</b>
3.1 ResNet5	7
3.2 EfficientNetB0	7
3.3 EfficientNetB3	7
3.4 Postupak treniranja	7
<b>4. EVALUACIJA IZGRAĐENIH METODA I ALGORITAMA</b>	<b>9</b>
4.1 Evaluacija RESNET50	9
4.2 Evaluacija EfficientNetB0	12
4.3 Evaluacija EfficientNetB3	14
4.4 Usporedba različitih strategije augmentacije podataka	18
4.4.1 Prva strategija augmentacije	18
4.4.2 Druga strategija augmentacije	18
4.5 Sažetak evaluacije i usporedba svih modela	19
<b>5. DEMONSTRACIJA RAD</b>	<b>21</b>
<b>6. ZAKLJUČAK</b>	<b>24</b>
<b>LITERATURA</b>	<b>25</b>
<b>POPIS SLIKA</b>	<b>26</b>
<b>POPIS TABLICA</b>	<b>27</b>
<b>PRILOZI</b>	<b>28</b>

## 1. UVOD

Koža je najveći organ ljudskog tijela i podložna je raznim bolestima koje mogu značajno utjecati na kvalitetu života. Precizna i pravovremena dijagnoza kožnih bolesti ključna je za uspješno liječenje i poboljšanje ishoda za pacijente. Međutim, zbog velike učestalosti i raznih oblika u kojima se kožne bolesti mogu javiti, pravilno dijagnosticiranje bolesti često zahtijeva brzo i pravovremeno iskustvo i intervenciju dermatologa, što nije uvijek moguće.

U suvremenoj medicini razvoj umjetne inteligencije, a posebice razvoj dubokog učenja, omogućuje napredak u analizi i automatskoj klasifikaciji medicinskih slika. Automatizirani sustavi temeljeni na dubokom učenju mogu poslužiti kao pomoćni alati liječnicima, smanjujući rizik od pogrešne dijagnoze i ubrzavajući proces donošenja odluka, osobito u područjima gdje je pristup dermatologima ograničen. Osim toga, takvi sustavi također imaju mogućnost otkrivanja malignih promjena kože, što ukazuje na hitnost intervencije medicinskih djelatnika.

Projekt se fokusira se na razvoj modela dubokog učenja za klasifikaciju kožnih bolesti na temelju slika, koristeći javno dostupne skupove podataka i napredne arhitekture konvolucijskih neuronskih mreža (CNN). Cilj je razviti sustav koji može pridonijeti bržoj, dostupnijoj i pouzdanijoj dijagnostici kožnih bolesti u kliničkoj praksi.

## 2. SKUP PODATAKA I ANALIZA

### 2.1 Opis skupa podataka

Za potreba ovog projekta korišten HAM10000 (Human Against Machine with 10000 training images), koji je jedan od najpoznatijih javno dostupnih skupa podataka za klasifikaciju kožnih bolesti. Skup sadrži ukupno 10015 dermatoskopskih slika različitih kožnih poremećaja. Svaka slika dolazi s pripadajućim metapodacima i oznakom klase kojoj pripada. [1] [2]


#### 2.1.1 Klase

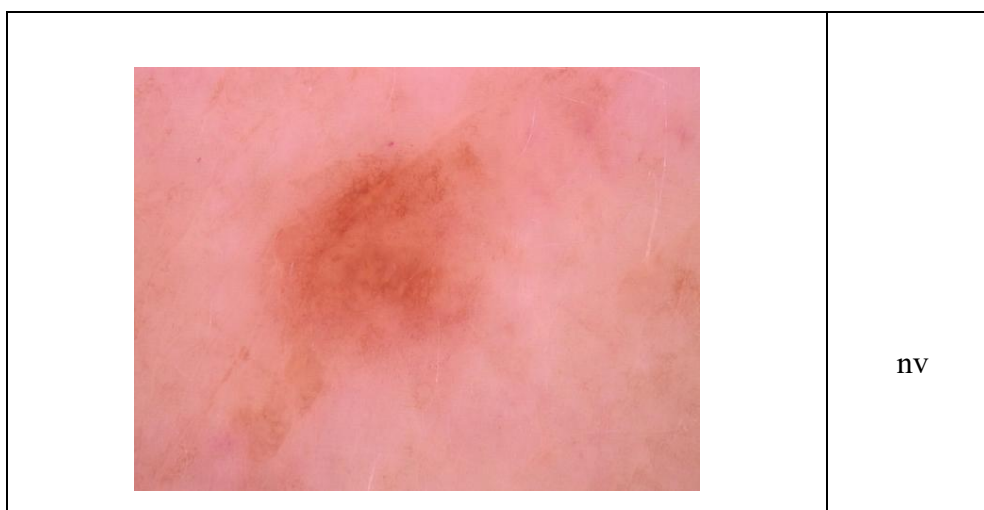
Svaka slika spomenutog skupa podataka pripada jednoj od 7 klasa:

- akiec: actinic keratoses and intraepithelial carcinoma (pre-malignant lesions)
- bcc: basal cell carcinoma (malignant)
- bkl: benign keratosis-like lesions (benign)
- df: dermatofibroma (benign)
- mel: melanoma (malignant)
- nv: melanocytic nevi (benign)
- vasc: vascular lesions (benign)

#### 2.1.2 Oznake i primjer podataka

Svaka slika označena je odgovarajućom oznakom klase (npr. “mel” za melanoma), čiji je primjer dan tablicom 2.1.

Slika (primjer)	Oznaka
	akiec



Tablica 2.1 Prikaz slika sa odgovarajućom oznakom klase

Uz slike dolazi i tablica metapodataka (HAM10000\_metadata.csv) koja za svaku sliku sadrži identifikator slike, oznaku klase, spol i dob pacijenta te lokalizaciju lezije na tijelu što je prikazano tablicom 2.2. [2]

image_id	dx	age	sex	localization
ISIC_0025030	bkl	80.0	male	scalp

Tablica 2.2 Primjer retka iz metapodataka

## 2.2 Podjela na skupove za trening i testiranje

Skup podataka podijeljen je na trening skup i testni skup u omjeru 80:20. Što znači da Trening skup sadrži 8012 slika, dok su 2003 slike korištene za test skup.

## 2.3 Rezolucija i format slika

Slike HAM10000 skupa podataka su u JPEG formatu i izvornoj rezoluciji 600x450 piksela. Prije ulaska u neuronsku mrežu, slike su dodatno skalirane na standardnu ulaznu veličinu arhitekture (npr. 224x224 piksela ili 300x300 piksela).

## 2.4 Eksplorativna analiza podataka

Analizom metapodataka utvrđeno je da je najzastupljenija klasa “nv” (melanocytic nevi, odnosno madeži), dok su klase “df” i “vasc” najmanje brojne. Također je primijećena značajna

neravnoteža među pojedinim klasama, kao i raspodjela po dobi i spolu pacijenata, što može dodatno utjecati na performanse modela. Prikaz broja slika po klasi dan je tablicom 2.3.

Klasa	Broj slika
nv	6705
mel	1113
bkl	1099
bcc	514
akiec	327
vasc	142
df	115

Tablica 2.3 Prikaz broja slika za pojedinu klasu

## 2.5 Predobrada i augmentacija podataka

Zbog izražene neravnoteže među klasama unutar skupa podataka HAM10000, kao i u cilju poboljšanja sposobnosti modela da generalizira na nove i nepoznate primjere, provedene su razne metode predobrade i augmentacije podataka. Ove metode bile su ključne za poboljšanje robusnosti modela, povećanje učinkovitosti učenja i smanjenje utjecaja klasne pristranosti.

U početnom koraku sve su slike skalirane na standardizirane dimenzije koje odgovaraju ulaznim zahtjevima odabranih dubokih arhitektura. Korišten format od  $224 \times 224$  piksela, što odgovara dimenzijama koje zahtijevaju modeli poput ResNet50 i EfficientNet-B0/B3. Ovisno o eksperimentu, slike su prilagođavane korištenjem različitih tehnika skaliranja: Resize za fiksno mijenjanje veličine, CenterCrop za izdvajanje centralnog dijela slike, te RandomResizedCrop za dodatnu varijaciju u prikazu i kompoziciji ulaza.

Kao standardni korak predobrade primijenjena je normalizacija piksela prema srednjim vrijednostima i standardnim devijacijama poznatog ImageNet skupa podataka (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). Time se omogućilo konzistentno ponašanje modela tijekom transfer learninga, s obzirom na to da su svi korišteni modeli prethodno trenirani upravo na ImageNet podacima.

U svrhu smanjenja overfittinga i povećanja varijabilnosti skupa podataka korištena je augmentacija slika. Primijenjene su dvije vrste augmentacijskih strategija.

Jednostavnija augmentacija koristila je osnovne transformacije iz `torchvision.transforms` modula. Uključivala je:

- Resize slike na standardne dimenzije
- `RandomHorizontalFlip` (nasumično horizontalno zrcaljenje)
- `RandomRotation` do 15 stupnjeva
- `ColorJitter` koji nasumično mijenja svjetlinu, kontrast i zasićenost slike

Ova jednostavnija augmentacija se pokazala posebno uspješnom u kombinaciji s modelom `EfficientNet-B0`, gdje je postignut najbolji ukupni rezultat na testnom skupu.

Naprednija augmentacija provedena je pomoću `Albumentations` biblioteke koja omogućuje kompleksnije i raznolikije transformacije. Među njima su bile:

- `RandomResizedCrop`, koji omogućuje skaliranje i slučajni izrez dijela slike
- Horizontal i Vertical Flip (zrcaljenja po horizontalnoj i vertikalnoj osi)
- Rotate do 30°
- `ColorJitter` sa širim rasponom intenziteta
- Afiine transformacije poput translacije i skaliranja
- `GaussianBlur`, kojim se uvodi kontrolirano zamućenje slike

Iako su ovakve tehnike generirale znatno veću varijabilnost unutar skupa podataka, uočeno je da preagresivne transformacije – osobito one koje značajno narušavaju anatomsku i vizualnu strukturu lezije – mogu imati negativan utjecaj na performanse modela. To se pokazalo kroz usporedbu rezultata evaluacije, gdje je jednostavnija augmentacija imala bolji učinak od kompleksnijih postupaka.

Što se tiče neravnoteže klasa – što je čest problem u medicinskim skupovima podataka – posebna pažnja posvećena je smanjenju utjecaja dominantnih klasa. Skup `HAM10000` sadržava izraženu dominaciju klase “melanocytic nevi” (nv), koja čini više od polovice svih uzoraka. Stoga su primijenjene sljedeće tehnike:

- Stratificirana podjela skupa na trening, validacijski i testni dio kako bi se očuvala klasna distribucija
- Težine u funkciji gubitka – korištenjem opcije `class_weight='balanced'` u `CrossEntropyLoss`, gdje su klasama dodijeljene težine obrnuto proporcionalne njihovoj učestalosti

- Nije korišten eksplicitni oversampling kako bi se izbjegla prenaučenosť modela na duplicirane primjere.

Zahvaljujući kombinaciji svih navedenih metoda predobrade i augmentacije, uspješno je povećana robusnost modela, smanjen rizik od prenaučenosťi i osigurana bolja generalizacija na nepoznate podatke bez potrebe za umjetnim povećavanjem skupa dupliciranjem uzoraka.



### **3. OPIS KORIŠTENIH METODA I ALGORITAMA**

U ovom projektu korištena su tri moderna modela dubokog učenja za klasifikaciju kožnih bolesti, i to ResNet50, EfficientNetB0 i EfficientNetB3. Svi modeli temelje se na konvolucijskim neuronskim mrežama unaprijed trenirani na velikom skupu prirodnih slika ImageNet, a za potrebe ovog zadatka dodatno su prilagođeni metodom transfer learninga. Ovim pristupom iskorištavaju se već naučene značajke modela, dok se završni slojevi posebno treniraju za prepoznavanje i klasifikaciju sedam različitih vrsta promjena na koži.[3] [5]

#### **3.1 ResNet5**

ResNet50 je duboka konvolucijska neuronska mreža (CNN) koja koristi rezidualne blokove za efikasno učenje bez gubitka točnosti pri povećanju dubine mreže. U ovom projektu, završni sloj originalnog modela zamijenjen je slojem sa sedam izlaza, koliko ima klasa bolesti u skupu podataka. Ključan korak pri treniranju je zamrzavanje svih slojeva osim novog klasifikacijskog sloja, čime se omogućuje brže učenje i smanjuje rizik od overfittinga.

#### **3.2 EfficientNetB0**

EfficientNetB0 uvodi princip skaliranja širine, dubine i rezolucije mreže na uravnotežen način. Ova arhitektura je laganija i brža u odnosu na ResNet, s manjim brojem parametara, ali zadržava visoku točnost. I ovdje je korišten transfer learning, završni sloj je prilagođen za klasifikaciju sedam klasa kožnih bolesti, dok su ostali slojevi inicijalno zamrznuti tijekom prvih faza treniranja.

#### **3.3 EfficientNetB3**

EfficientNetB3 predstavlja unaprijeđenu verziju B0 modela te koristi dublju i širu mrežu uz veće ulazne rezolucije slika, što često dovodi do još bolje točnosti, ali zahtijeva i veće računalne resurse. Model je treniran metodom transfer learninga, a završni sloj je prilagođen za klasifikaciju sedam različitih klasa bolesti kože.

#### **3.4 Postupak treniranja**

Za sva tri modela korišten je isti proces treniranja. Kao optimizator korišten je Adam s početnim learning rateom od  $1e-4$ , dok je funkcija gubitka bila Cross-Entropy Loss. Batch size je iznosio 32 slike po iteraciji. Prvo su trenirani samo novi (klasifikacijski) slojevi uz zamrznute ostatke modela (feature extractor), nakon čega se po potrebi pristupalo finom podešavanju

(fine-tuning) cijele mreže. Za smanjenje overfittinga primijenjene su različite tehnike, uključujući augmentaciju podataka, dropout, early stopping i weight decay. Tijekom treniranja, modeli su redovito evaluirani na validacijskom skupu.

Posebno kod treniranja EfficientNet modela, dodatna pažnja posvećena je postupku predobrade slika. Slike su, prije ulaska u model, skalirane na odgovarajuću veličinu (224x224 piksela za EfficientNetB0 i 300x300 piksela za EfficientNetB3), a pikselske vrijednosti normalizirane su prema statistikama ImageNet skupa podataka. Uz standardne augmentacijske tehnike (rotacije, zrcaljenje, promjene kontrasta i svjetline), korišten je i specifičan PyTorch transform pipeline prilagođen preporukama za EfficientNet arhitekture, kako bi se maksimalno iskoristile sposobnosti ovih modela. Ova prilagođena predobrada omogućila je modelima bolju generalizaciju na testnim podacima i povećala ukupnu točnost klasifikacije.

## 4. EVALUACIJA IZGRAĐENIH METODA I ALGORITAMA

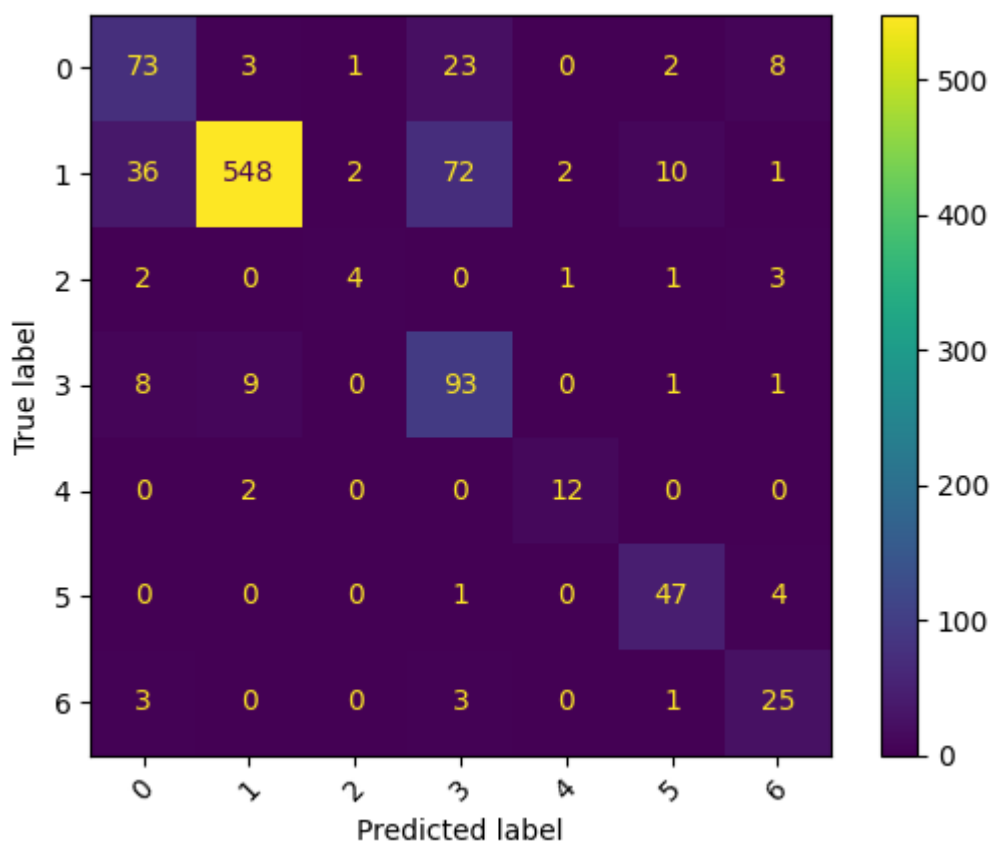
### 4.1 Evaluacija RESNET50

Model ResNet50 treniran je na HAM10000 skupu koristeći osnovnu augmentaciju podataka koja uključuje promjenu veličine, horizontalno preslikavanje, rotaciju i jitter. Podaci su podijeljeni u 80% za treniranje, 10% za validaciju i 10% za testiranje, uz zadržavanje proporcionalne zastupljenosti klasa.

Metrika	Vrijednost
Točnost	0.8004
F1-score	0.7003
Recall	0.7452
Preciznost	0.6832
Test loss	0.9332

Tablica 4.1 Mjerne vrijednosti performansi modela ResNet50 na testnom skupu

Dobivena točnost ukazuje na solidne performanse, no niži F1-score i preciznost pokazuju da model ima poteškoća s konzistentno točnim klasifikacijama među svim klasama. Relativno visok recall sugerira da model dobro prepoznaje većinu pozitivnih slučajeva, što je važno za medicinske aplikacije, ali niža preciznost upućuje na moguću pojavu lažno pozitivnih predikcija. Viši testni gubitak u odnosu na druge testirane modele ukazuje i na slabiju sposobnost generalizacije.



Slika 4.1 Matrica zabune za ResNet50

Kao nadopuna kvantitativnim metrikama, provedena je detaljna analiza performansi ResNet50 modela pomoću matrice zabune i klasifikacijskog izvještaja. Ove analize omogućuju uvid u ponašanje modela po klasama i identifikaciju konkretnih izazova u klasifikaciji.

Iz matrice zabune vidljivo je da model najuspješnije klasificira klasu „nv” (melanocytic nevi), koja je ujedno i najzastupljenija u skupu, s ukupno 548 točno klasificiranih slučajeva. S druge strane, model pokazuje izražene poteškoće u razlikovanju klasa „bkl” (benign keratosis-like lesions) i „mel” (melanoma), što je osobito važno u kliničkom kontekstu jer su te dvije dijagnoze često vizualno slične, ali klinički vrlo različite. Klasa „mel” je često pogrešno klasificirana kao „bkl” (23 slučaja) i kao „nv” (10 slučajeva), što ukazuje na potrebu za dodatnim poboljšanjima u prepoznavanju malignih lezija.

	precision	recall	f1-score	support
bkl	0.60	0.66	0.63	110
nv	0.98	0.82	0.89	671
df	0.57	0.36	0.44	11
mel	0.48	0.83	0.61	112
vasc	0.80	0.86	0.83	14
bcc	0.76	0.90	0.82	52
akiec	0.60	0.78	0.68	32
accuracy			0.80	1002
macro avg	0.68	0.75	0.70	1002
weighted avg	0.85	0.80	0.81	1002

Slika 4.2 Klasifikacijski izvještaj modela ResNet50 po klasama na testnom skupu

Klasifikacijski izvještaj na slici prikazuje vrijednosti preciznosti, odziva i F1-mjere po svakoj pojedinačnoj klasi na testnom skupu od 1002 uzorka. Također su uključeni makro i ponderirani prosjeci metrika, kao i ukupna točnost modela.

Najbolji rezultati postignuti su za klasu „nv” (melanocytic nevi), gdje model postiže preciznost od 0.98, odziv od 0.82 i F1-mjeru od 0.89, što je rezultat velikog broja primjera i lako prepoznatljivih karakteristika ove klase. Tako visoki rezultati upućuju na to da model iznimno pouzdano klasificira slike ove najzastupljenije klase.

Klasa „mel” (melanoma) ima visok odziv od 0.83, ali znatno nižu preciznost (0.48), što znači da model relativno uspješno prepoznaje melanome kada su prisutni, no često ih pogrešno dodjeljuje i kada to nije slučaj (lažno pozitivni slučajevi). Ovo je posebno važno u medicinskom kontekstu jer lažno pozitivne dijagnoze mogu dovesti do nepotrebne zabrinutosti ili intervencija.

Klase „bcc” (basal cell carcinoma) i „vasc” (vascular lesions) također pokazuju visoke vrijednosti F1-mjere (0.82 i 0.83), što ukazuje na dobar balans između preciznosti i odziva. Iako su ove klase manje zastupljene, model ih klasificira relativno pouzdano, što sugerira da imaju specifične vizualne obrasce koje model lako prepoznaje.

S druge strane, klase s manje podataka, poput „df” (dermatofibroma) i „akiec” (actinic keratoses), pokazuju znatno slabije rezultate. „df” ima najniži F1-score (0.44), što je očekivano

s obzirom na to da ima samo 11 primjera u testnom skupu i vrlo slične karakteristike drugim benignim lezijama.

Zaključno, iako ResNet50 postiže prihvatljive rezultate, pokazuje određena ograničenja u preciznosti i pouzdanosti, što ga čini manje pogodnim za praktičnu primjenu u kontekstu dijagnostike kožnih bolesti.

## 4. 2 Evaluacija EfficientNetB0

Model EfficientNetB0 treniran je na istom skupu podataka i pod istim uvjetima kao i prethodni modeli, koristeći osnovnu strategiju augmentacije (resize, horizontalni flip, rotacija, jitter, normalizacija). Ova arhitektura je odabrana zbog svoje poznate efikasnosti i uravnoteženja između broja parametara i performansi.

Metrika	Vrijednost
Točnost	0.8533
F1-score	0.7435
Recall	0.7906
Preciznost	0.7094
Test loss	0.7978

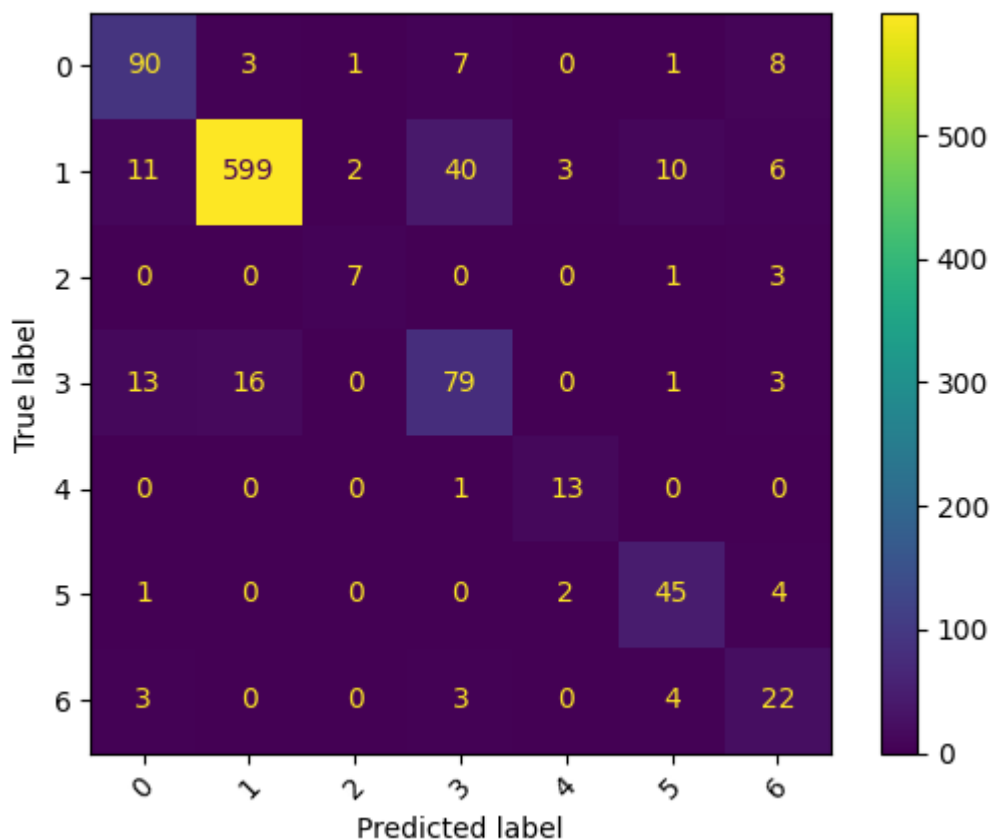
Tablica 4.2 Mjerne vrijednosti performansi modela EfficientNetB0 na testnom skupu

Ostvarena točnost od 85,3% potvrđuje da model ima visoku sposobnost klasifikacije medicinskih slika na temelju vizualnih obilježja. F1-mjera od 74,35% ukazuje na dobru ravnotežu između preciznosti i odziva na razini svih klasa, a osobito visoki recall (79%) sugerira da je model uspješan u otkrivanju većine pozitivnih slučajeva, što je osobito važno u kontekstu ranog otkrivanja bolesti.

Usporedno s ResNet50 modelom, EfficientNetB0 pokazuje niži testni gubitak (0.7978), što ukazuje na bolju sposobnost generalizacije i manji broj pogrešaka pri klasifikaciji neviđenih uzoraka.

Dobiveni rezultati potvrđuju da je EfficientNetB0 vrlo pogodan za zadatak klasifikacije kožnih bolesti te predstavlja značajan napredak u odnosu na klasičnije arhitekture. Posebno se ističe svojom efikasnošću u kombinaciji s visokim performansama, čime se pokazuje kao dobar

kandidat za daljnji razvoj i potencijalnu primjenu u medicinskoj praksi, uz dodatna sigurnosna mjerenja i verifikaciju.



Slika 4.3 Matrica zabune za EfficientNetB0

Matrica zabune pokazuje da model najtočnije klasificira klasu „nv” (melanocytic nevi), s 599 točno klasificiranih primjera od ukupno 671. Također, klasa „bkl” (benign keratosis-like lesions) je većinom ispravno prepoznata, s 90 točnih predikcija od 110. Najveći broj pogrešnih klasifikacija vidljiv je kod klase „mel” (melanoma), koja se često zamjenjuje s klasama „nv” i „bkl”. Konkretno, model je klasu „mel” u 13 slučajeva klasificirao kao „bkl”, a u 16 kao „nv”, što ukazuje na poteškoće u razlikovanju vizualno sličnih, ali klinički različitih lezija. Klasa „akiec” (actinic keratoses) također pokazuje nižu točnost, pri čemu se model često dvoumi između nje i klasa „bkl” i „bcc”.

	precision	recall	f1-score	support
bkl	0.76	0.82	0.79	110
nv	0.97	0.89	0.93	671
df	0.70	0.64	0.67	11
mel	0.61	0.71	0.65	112
vasc	0.72	0.93	0.81	14
bcc	0.73	0.87	0.79	52
akiec	0.48	0.69	0.56	32
accuracy			0.85	1002
macro avg	0.71	0.79	0.74	1002
weighted avg	0.87	0.85	0.86	1002

Slika 4.4 Klasifikacijski izvještaj modela EfficientNetB0 po klasama na testnom skupu

Klasifikacijski izvještaj dodatno potvrđuje ove nalaze. Klasa „nv” postiže izuzetno visoke vrijednosti: preciznost 0.97, odziv 0.89 i F1-mjera 0.93, što odražava dominantnu prisutnost te klase u skupu i njenu prepoznatljivost. Klase „bkl“, „bcc“ i „vasc“ također pokazuju dobre rezultate, s F1-mjerama od 0.79, 0.79 i 0.81, čime se potvrđuje pouzdanost modela u prepoznavanju tih dijagnoza. S druge strane, klasa „mel” ima nešto nižu preciznost (0.61), ali relativno dobar odziv (0.71), što znači da model često prepoznaje melanome, no u određenim slučajevima daje lažno pozitivne predikcije. Manje zastupljene klase, poput „df“ i „akiec“, i dalje imaju niže rezultate, ali ipak bolje nego kod prethodnih modela – primjerice, klasa „df“ ima F1-mjeru od 0.67 (u usporedbi s 0.44 kod ResNet50).

Ukupno gledano, model EfficientNetB0 ostvario je točnost od 85% na testnom skupu, dok su makro prosjeci iznosili: preciznost 0.71, odziv 0.79 i F1-mjera 0.74. Ponderirani prosjeci su još viši, s F1-mjerom od 0.86, što je rezultat izrazito dobre klasifikacije većinskih klasa. Ove vrijednosti ukazuju na uravnotežene i stabilne performanse modela, te predstavljaju značajan napredak u odnosu na prethodno testirane arhitekture.

### 4.3 Evaluacija EfficientNetB3

Treći testirani model bio je EfficientNetB3, veća i dublja verzija arhitekture EfficientNet, s više parametara i slojeva u odnosu na B0. Cilj je bio ispitati hoće li dodatna dubina i kapacitet modela rezultirati boljim performansama u klasifikaciji kožnih bolesti.



Model je treniran pod istim uvjetima kao i prethodni, koristeći osnovnu strategiju augmentacije te identičnu podjelu podataka.

Metrika	Vrijednost
Točnost	0.8603
F1-score	0.7633
Recall	0.8217
Preciznost	0.072
Test loss	0.5810

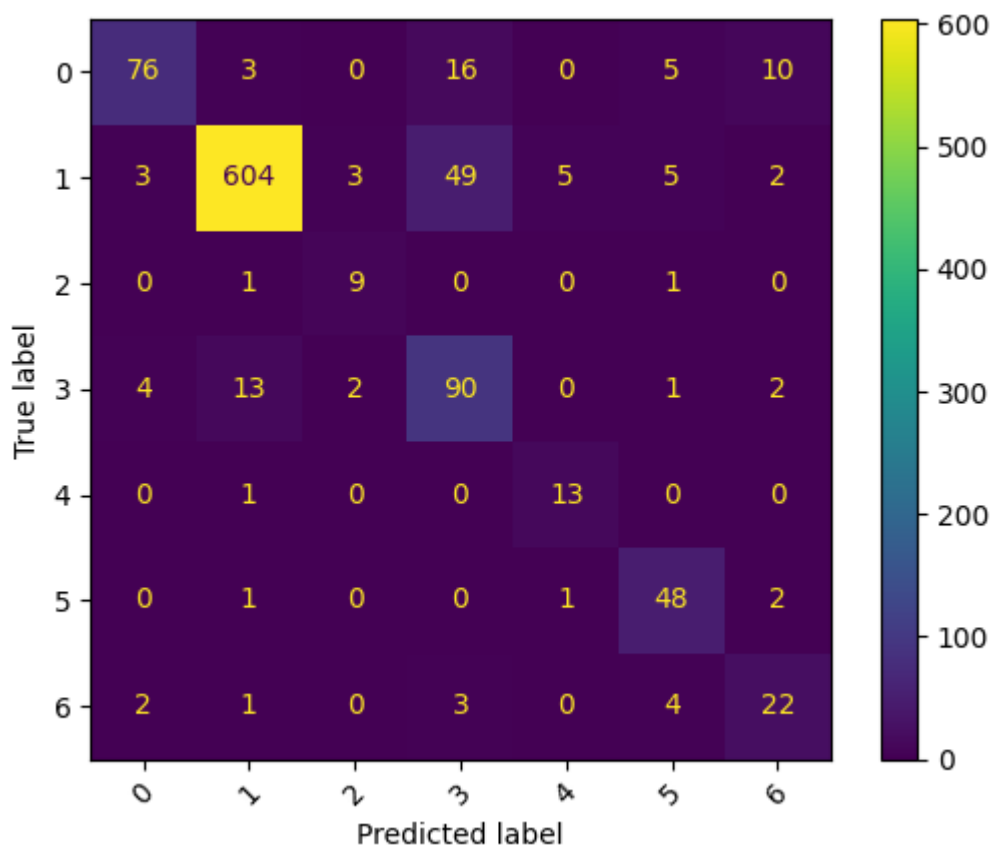
Tablica 4.3 Mjerne vrijednosti performansi modela EfficientNetB3 na testnom skupu

Dobiveni rezultati pokazuju da EfficientNetB3, nakon dodatnih optimizacija, ostvaruje najbolje ukupne performanse među svim testiranim modelima. Postignute su sljedeće vrijednosti: točnost od 86,03%, F1-mjera od 76,33%, odziv od 82,17% i preciznost od 72,68%. Ove vrijednosti nadmašuju rezultate dobivene s modelima ResNet50 i EfficientNetB0, čime se EfficientNetB3 potvrđuje kao najuspješnija arhitektura u okviru ovog rada.

Posebno visoka vrijednost recall-a (82,17%) naglašava sposobnost modela da detektira većinu pozitivnih slučajeva, što je ključno u medicinskim primjenama, gdje je izbjegavanje lažno negativnih predikcija od presudnog značaja. Također, poboljšanje F1-mjere ukazuje na uravnotežen odnos između preciznosti i odziva na razini svih klasa.

Međutim, takve performanse dolaze uz značajnu cijenu u pogledu računalnih resursa i vremena treniranja. U usporedbi s EfficientNetB0, koji je poznat po efikasnosti i brzom treniranju, model EfficientNetB3 zahtijevao je višestruko duže vrijeme za treniranje zbog većeg broja slojeva i parametara. Trening je trajao znatno duže i bio je znatno zahtjevniji po pitanju memorije GPU-a i ukupne računalne snage. U odnosu na ResNet50, razlika je bila još izraženija – dok se ResNet50 trenirao u razumnom vremenskom okviru, EfficientNetB3 je zahtijevao i višestruko dulje epohe kako bi postigao stabilnu konvergenciju.

Unatoč toj složenosti i opterećenju, dodatno uloženi resursi i vrijeme treniranja pokazali su se opravdanim, s obzirom na jasno mjerljivo poboljšanje klasifikacijskih performansi. U konačnici, EfficientNetB3 se nameće kao najučinkovitiji i najpouzdaniji model unutar ovog istraživanja, sposoban za visoko točnu klasifikaciju medicinskih slika, i kao takav predstavlja najboljeg kandidata za daljnju primjenu u kliničkom kontekstu, uz uvjet dostupnosti odgovarajuće računalne infrastrukture.



Slika 4.5 Matrica zabune za EfficientNetB3

Za model EfficientNetB3 provedena je dodatna analiza po klasama korištenjem matrice zabune i klasifikacijskog izvještaja. Cilj je bio bolje razumjeti u kojim slučajevima model postiže dobre rezultate, a gdje još uvijek dolazi do pogrešaka.

Matrica zabune pokazuje da model najtočnije klasificira klasu „nv”, s ukupno 604 točne klasifikacije od 671, što je u skladu s ranijim rezultatima. Također, klase „bcc” i „vasc” prepoznate su s visokom točnošću – primjerice, „vasc” je gotovo savršeno klasificiran, sa samo jednom pogrešnom predikcijom. S druge strane, značajan broj pogrešaka javlja se u klasifikaciji klasa „bkl”, „mel” i „akiec”. Klasa „mel” najčešće se pogrešno klasificira kao „nv” (13 slučajeva) i „bkl” (4 slučajeva), što se poklapa s očekivanjima s obzirom na vizualnu sličnost tih lezija.

	precision	recall	f1-score	support
bkl	0.89	0.69	0.78	110
nv	0.97	0.90	0.93	671
df	0.64	0.82	0.72	11
mel	0.57	0.80	0.67	112
vasc	0.68	0.93	0.79	14
bcc	0.75	0.92	0.83	52
akiec	0.58	0.69	0.63	32
accuracy			0.86	1002
macro avg	0.73	0.82	0.76	1002
weighted avg	0.88	0.86	0.87	1002

Slika 4.6 Klasifikacijski izvještaj modela EfficientNetB3 po klasama na testnom skupu

Klasifikacijski izvještaj prikazuje da model EfficientNetB3 ostvaruje izuzetno uravnotežene performanse na razini pojedinačnih klasa. Klasa „nv” (melanocytic nevi) postiže najbolji rezultat, s preciznošću od 0.97, recall-om od 0.90 i F1-mjerom od 0.93, što je rezultat velikog broja primjera i jasnih vizualnih obrazaca koji ovu klasu čine lako prepoznatljivom. Dobre rezultate postižu i klase „bcc” i „vasc”, koje ostvaruju F1-mjere od 0.83 i 0.79 uz visoke vrijednosti recall-a (0.92 i 0.93), što ukazuje na visoku pouzdanost modela u detekciji tih lezija. Klasa „df” također ostvaruje stabilnu F1-mjeru od 0.72, unatoč malom broju primjera. Posebnu pažnju zaslužuje klasa „mel” (melanoma), koja postiže F1-mjeru od 0.67 i recall od 0.80, što znači da model relativno uspješno prepoznaje većinu melanoma. Ovo je izuzetno važno za praktičnu primjenu modela u medicinskom kontekstu, budući da propuštanje malignih lezija može imati ozbiljne posljedice. S druge strane, najslabiji rezultat ostvaren je kod klase „akiec”, s F1-mjerom od 0.63, što je još uvijek u prihvatljivom rasponu, ali ukazuje na potrebu za dodatnim unapređenjem detekcije manje zastupljenih klasa.

Ukupna točnost modela iznosi 86%, dok makro prosjeci iznose: preciznost 0.73, recall 0.82 i F1-mjera 0.76. Ponderirani prosjeci su još viši, s F1-mjerom od 0.87, što potvrđuje da model pouzdanije klasificira većinski zastupljene klase, ali ostvaruje i značajan napredak kod manjih i kompleksnijih klasa.

Ovi rezultati dodatno potvrđuju da EfficientNetB3, u svojoj optimiziranoj konfiguraciji, pruža najuravnoteženije i najpouzdanije performanse od svih testiranih modela, čime se opravdava njegov odabir za primjenu u zadatku klasifikacije kožnih bolesti.

#### 4.4 Usporedba različitih strategije augmentacije podataka

Budući da se model EfficientNetB0 tijekom eksperimenta pokazao kao najbolji omjer između klasifikacijske točnosti i računalne učinkovitosti, dodatno su ispitane dvije različite strategije augmentacije podataka s ciljem procjene njihovog utjecaja na performanse modela. U oba slučaja korišten je isti model, isti skup podataka te identična podjela na trening, validacijski i testni dio.

Metrika	Prva strategija augmentacije (Albumentatioons)	Druga strategija augmentacije (torchvision.transforms)
Točnost	0.8174	0.8533
F1-score	0.6918	0.7435
Recall	0.7242	0.7906
Preciznost	0.6685	0.7094
Test loss	0.8047	0.7978

Tablica 4.4 Učinak različitih strategija augmentacije na performanse modela

##### 4.4.1 Prva strategija augmentacije

Prva strategija koristila je biblioteku Albumentations, koja omogućuje složenije i raznovrsnije transformacije. Ova augmentacija uključivala je slučajno kadriranje (crop), horizontalno i vertikalno preslikavanje, rotacije, jitter, affine transformacije i Gaussian blur. Cilj je bio stvoriti što raznovrsniji skup podataka koji potencijalno bolje generalizira.

Iako se očekivalo da će složenija augmentacija poboljšati performanse, rezultati su pokazali suprotno – model je ostvario niže metrike u odnosu na jednostavniju varijantu augmentacije.

##### 4.4.2 Druga strategija augmentacije

Druga strategija oslanjala se na osnovne transformacije dostupne u PyTorch biblioteci torchvision.transforms, koje uključuju promjenu veličine slike, horizontalno preslikavanje,

rotaciju do 15 stupnjeva te lagani jitter. Ova metoda je znatno jednostavnija, ali rezultati pokazuju da je bila uspješnija u ovom konkretnom slučaju.

Dobiveni rezultati ukazuju na to da kompleksnija augmentacija ne garantira bolje performanse, barem ne u svakom kontekstu. U ovom slučaju, jednostavnije transformacije rezultirale su boljim F1-scoreom, višom preciznošću i boljom točnošću, dok je složenija augmentacija mogla unijeti preveliku varijabilnost koja otežava učenje stabilnih reprezentacija. Ovo naglašava važnost prilagodbe augmentacijskih strategija specifičnostima problema, umjesto automatskog uvođenja kompleksnosti.

#### **4.5 Sažetak evaluacije i usporedba svih modela**

EfficientNetB0 se pokazao kao najučinkovitiji model u sklopu ovog istraživanja, s najvišom ukupnom točnošću (85,33%), najboljim F1-scoreom (74,35%) te najnižim testnim gubitkom (0,7978). Model je demonstrirao visoku stabilnost tijekom treniranja i vrlo dobru generalizaciju na testnom skupu, čime se istaknuo kao najsnažnija arhitektura u kontekstu zadanog problema. Unatoč većem broju slojeva, EfficientNetB3 nije uspio nadmašiti rezultate postignute verzijom B0. Iako je B3 model ostvario visoku točnost i uravnotežene metrike, njegov viši testni gubitak (0,8756) i niži F1-score (72,95%) sugeriraju da dodatna složenost nije donijela značajno poboljšanje performansi, a može čak implicirati i manju otpornost na prenaučenosť.

S druge strane, ResNet50, iako pouzdana i dobro poznata arhitektura, pokazao je najslabije rezultate među testiranim modelima. Postignuta točnost od 80,04% te F1-score od 70,03% ukazuju na solidne, ali osjetno slabije performanse u usporedbi s EfficientNet varijantama. Najviši testni gubitak (0,9332) među svim modelima dodatno potvrđuje manju sposobnost generalizacije ove arhitekture na složenom i neuravnoteženom medicinskom skupu podataka.

Važno je napomenuti da veća arhitektonska složenost ne garantira automatski i bolje rezultate. Ovaj zaključak potvrđuje usporedba između EfficientNetB0 i B3, gdje je jednostavnija i optimiziranija varijanta pokazala bolji ukupni učinak. Slično tome, provedeni testovi s različitim vrstama augmentacije podataka pokazali su da umjerena i ciljana augmentacija (osnovne transformacije poput horizontalnog preslikavanja i rotacije) može rezultirati boljim performansama od kompleksnijih strategija koje uključuju napredne kombinacije transformacija. Agresivnija augmentacija u nekim je slučajevima dovela do prevelike varijabilnosti podataka i smanjene preciznosti modela.

Na temelju svih analiziranih metrika, može se zaključiti da EfficientNetB0 pruža optimalan omjer između točnosti, efikasnosti, stabilnosti treniranja i opće primjenjivosti. U okviru ovog rada, ovaj model se nametnuo kao najpouzdaniji kandidat za zadatak klasifikacije medicinskih slika kožnih lezija, a njegova primjena može poslužiti kao temelj za budući razvoj kliničkih dijagnostičkih alata temeljenih na umjetnoj inteligenciji.

## 5. DEMONSTRACIJA RADA

a potrebe praktične demonstracije razvijen je interaktivni sučajni alat korištenjem biblioteke Streamlit, koji omogućuje jednostavnu i vizualno preglednu upotrebu modela za klasifikaciju kožnih bolesti [6]. Aplikacija je osmišljena kao prototip dijagnostičkog alata temeljenog na dubokom učenju, koji omogućuje korisniku da učitava sliku kožne lezije, odabere jedan od prethodno treniranih modela i dobije predikciju dijagnoze uz pripadajuću razinu sigurnosti.

U aplikaciji su dostupni svi trenirani modeli iz eksperimentalnog dijela rada, uključujući:

- EfficientNet-B0
- EfficientNet-B0 (s drugačijim preprocesiranjem)
- EfficientNet-B3
- ResNet50

Nakon što korisnik učitava sliku, model provodi odgovarajuću predobradu slike, izvršava klasifikaciju, te vraća naziv predviđene klase bolesti (npr. *melanoma*, *nevus*, *basal cell carcinoma*, itd.) zajedno s postotkom sigurnosti predikcije. Rezultat se dodatno prikazuje u obliku stupčastog grafa koji vizualizira vjerojatnosti za sve dostupne klase. Time se korisniku omogućuje bolji uvid u raspodjelu predikcijskih vrijednosti i potencijalnu nesigurnost modela. Slike 5.1 I 5.2 jasno prikazuju izgled korisničkog sučelja sa učitavanjem slike u aplikaciju te način na koji se rezultati predikcije i vjerojatnosti za pojedine klase vizualno prezentiraju korisniku.

## Skin Disease Diagnosis

Upload a photo of a skin lesion and select the deep learning model for diagnosis. The predicted class and confidence will be shown.

Choose a model:

EfficientNet-B0

Upload a skin image (JPG, PNG):



Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

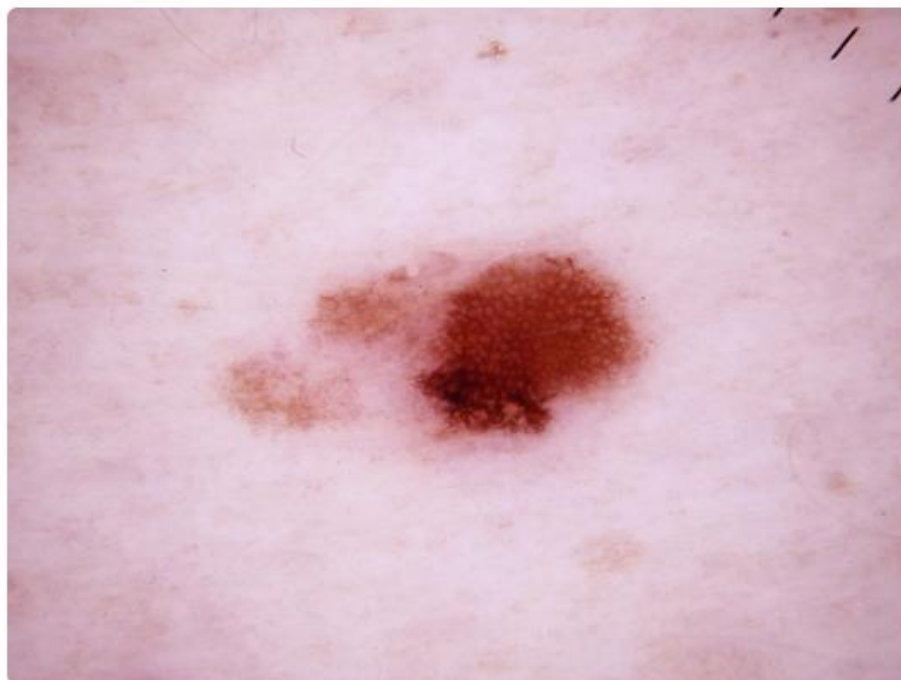
Browse files



ISIC\_0024313.jpg 242.1KB



The `use_column_width` parameter has been deprecated and will be removed in a future release. Please utilize the `use_container_width` parameter instead.

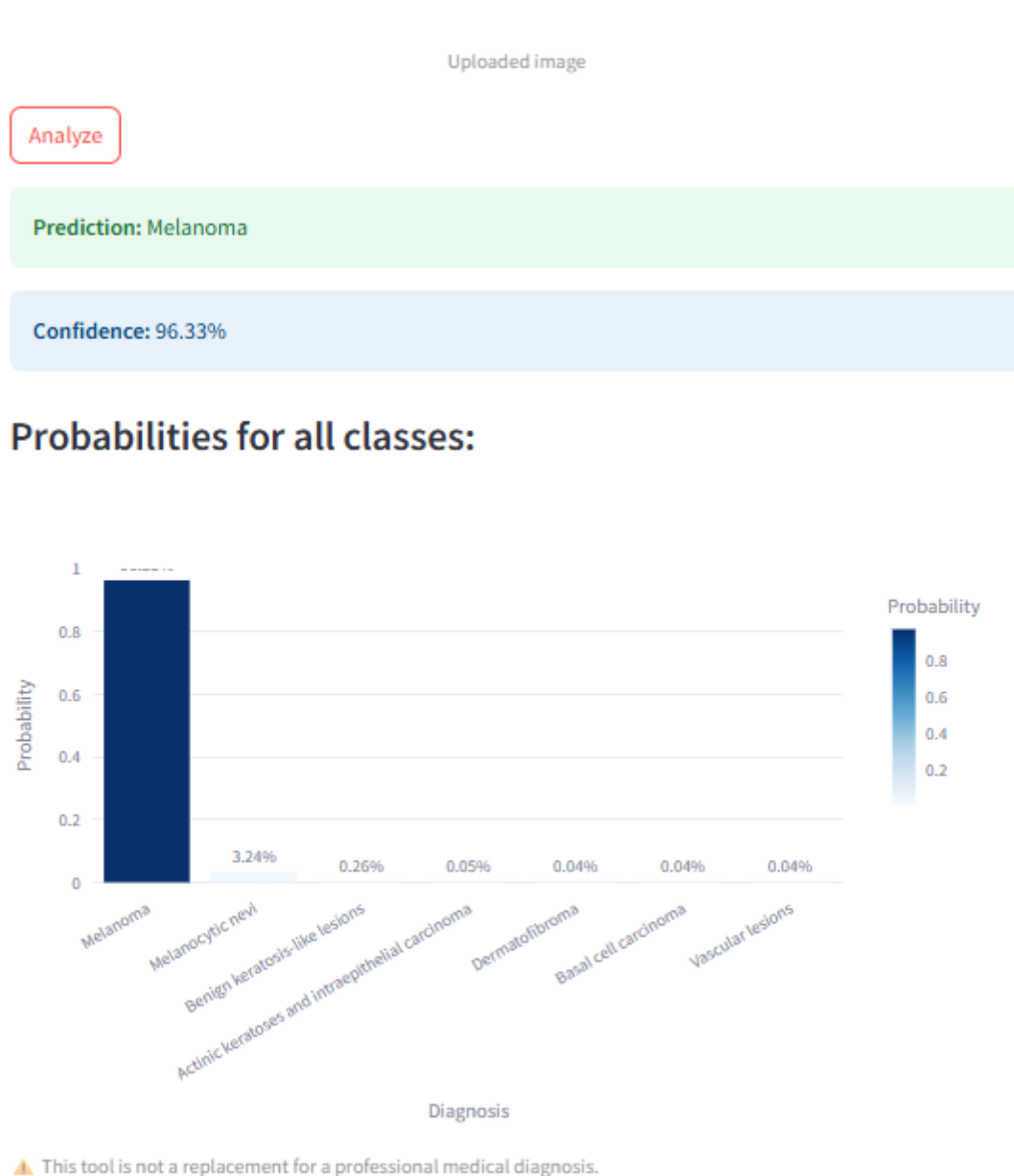


Uploaded image

Analyze

Slika 5.1 Sučelje Streamlit aplikacije za dijagnostiku kožnih bolesti





Slika 5.2 Rezultati analize i vizualizacija vjerojatnosti po klasama

## 6. ZAKLJUČAK

U ovom radu razvijen je i evaluiran sustav temeljen na dubokom učenju za klasifikaciju kožnih bolesti korištenjem medicinskih slika iz javno dostupnog skupa podataka HAM10000. Cilj je bio ispitati učinkovitost različitih konvolucijskih neuronskih mreža u prepoznavanju sedam tipova lezija kože, pri čemu je poseban naglasak stavljen na točnost, F1-mjeru, odziv i preciznost, kao i na sigurnost primjene u medicinskom kontekstu.

Rezultati su pokazali da EfficientNetB3, nakon dodatnih optimizacija, ostvaruje najbolje ukupne performanse s točnošću od 86,03%, F1-mjerom od 76,33% i recallom od 82,17%, što ga čini najpreciznijim i najpouzdanijim modelom unutar ovog istraživanja. Ipak, postizanje ovakvih rezultata zahtijevalo je znatno dulje vrijeme treniranja i veće računalne resurse u odnosu na druge arhitekture. S druge strane, model EfficientNetB0 istaknuo se kao najbolji kompromis između točnosti i računalne učinkovitosti, s vrlo konkurentnim rezultatima (točnost 85,3%, F1-mjera 74,3%) i značajno manjim zahtjevima za treniranje. Ova arhitektura se pokazala optimalnom za evaluaciju učinka različitih augmentacijskih strategija te kao dobra osnova za sustave koji trebaju brzu obradu i manju potrošnju resursa. Usporedba augmentacijskih strategija dodatno je pokazala da umjerena augmentacija donosi stabilnije rezultate od agresivnijih pristupa, što je važan zaključak za rad s osjetljivim medicinskim podacima.

Za demonstraciju praktične primjene razvijena je interaktivna aplikacija u Streamlitu, koja korisniku omogućuje učitavanje slike kože, odabir modela i prikaz predikcije s pripadajućom sigurnošću. Time je pokazano da je moguće razviti funkcionalan prototip dijagnostičkog alata temeljen na dubokom učenju, koji može poslužiti kao podrška medicinskom osoblju u preliminarnoj procjeni kožnih lezija.

Iako su postignuti rezultati obećavajući, važno je naglasiti da se ovakvi sustavi ne mogu koristiti kao zamjena za stručnu medicinsku dijagnozu. Za stvarnu kliničku primjenu bilo bi potrebno provesti dodatnu validaciju na većim i raznolikijim skupovima podataka, uvođenje sustava za procjenu pouzdanosti modela, te integraciju s drugim kliničkim informacijama.

Zaključno, ovaj rad pokazuje da duboko učenje predstavlja snažan alat za automatizaciju dijagnostike dermatoloških bolesti, te otvara prostor za daljnji razvoj i potencijalnu implementaciju u kliničkom okruženju, uz odgovarajuće kontrole i multidisciplinarni pristup.

## LITERATURA

- [1] Dhruv1234. (n.d.). *HAM10000 - Skin Disease Classification* [Notebook]. Kaggle.  
<https://www.kaggle.com/code/dhruv1234/ham10000-skin-disease-classification/notebook>
- [2] K. Mader, *Skin Cancer MNIST: HAM10000*, Kaggle. [Online]. Dostupno:  
<https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>.
- [3] ScienceDirect. (2024). *Enhancing skin lesion classification with advanced deep learning*.  
<https://www.sciencedirect.com/science/article/abs/pii/S0147027224000187>
- [4] Hall, M. X. (n.d.). *DL\_medical\_imaging\_cle-ai/HAM10000\_skin\_lesion\_classifier* [Notebook]. GitHub. [https://github.com/hallmx/DL\\_medical\\_imaging\\_cle-ai/tree/master](https://github.com/hallmx/DL_medical_imaging_cle-ai/tree/master)
- [5] Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmanan, K. (2022). *Skin lesion classification of dermoscopic images using machine learning and convolutional neural network*. *Scientific Reports*  
<https://www.nature.com/articles/s41598-022-22644-9>
- [6] Streamlit Documentation. <https://docs.streamlit.io/>

## POPIS SLIKA

Slika 4.1 Matrica zabune za ResNet50 .....	10
Slika 4.2 Klasifikacijski izvještaj modela ResNet50 po klasama na testnom skupu .....	11
Slika 4.3 Matrica zabune za EfficientNetB0.....	13
Slika 4.4 Klasifikacijski izvještaj modela EfficientNetB0 po klasama na testnom skupu .....	14
Slika 4.5 Matrica zabune za EfficientNetB3.....	16
Slika 4.6 Klasifikacijski izvještaj modela EfficientNetB3 po klasama na testnom skupu .....	17
Slika 5.1 Sučelje Streamlit aplikacije za dijagnostiku kožnih bolesti .....	22
Slika 5.2 Rezultati analize i vizualizacija vjerojatnosti po klasama .....	23

## POPIS TABLICA

Tablica 2.1 Prikaz slika sa odgovarajućom oznakom klase .....	2
Tablica 2.2 Primjer retka iz metapodataka .....	3
Tablica 2.3 Prikaz broja slika za pojedinu klasu .....	4
Tablica 4.1 Mjerne vrijednosti performansi modela ResNet50 na testnom skupu .....	9
Tablica 4.2 Mjerne vrijednosti performansi modela EfficientNetB0 na testnom skupu .....	12
Tablica 4.3 Mjerne vrijednosti performansi modela EfficientNetB3 na testnom skupu .....	15
Tablica 4.4 Učinak različitih strategija augmentacije na performanse modela .....	18

## PRILOZI

Cijeli izvorni kod korišten u ovom projektu, uključujući skripte za obradu podataka, treniranje modela i implementaciju Streamlit aplikacije, dostupan je u sljedećem GitHub repozitoriju:

<https://github.com/Samke004/Skin-Disease-Classification>

Repozitorij sadrži:

- Skripte za predobradu i augmentaciju podataka
- Jupyter bilježnice za treniranje i evaluaciju modela
- Kôd za Streamlit aplikaciju