



UNIVERZITET U SARAJEVU
ELEKTROTEHNIČKI FAKULTET
ODSJEK ZA RAČUNARSTVO I INFORMATIKU

Klasifikacija medicinskih slika ResNet BreastMNIST

Projekat iz predmeta
VJEŠTAČKA INTELIGENCIJA

Supervizor:

Efendić Emina

Studenti:

Čajić Amina 19218

Hodžić Una 19044

Sadržaj

1 Uvod i opis problema.....	5
1.1 Osnovni pojmovi.....	5
1.1.1 Osnovni tehnički pojmovi.....	5
1.1.1 Osnovni medicinski pojmovi.....	7
1.2 Koristi rješavanja problema u kliničkoj praksi.....	7
1.3 Pregled postojećih dataset-ova.....	8
2 Pregled stanja u oblasti.....	9
3 Izbor, analiza i pretprocesiranje dataset-a.....	11
3.1 Izvor dataset-a.....	11
3.2 Format u kojem je dostupan i način za preuzimanje i korištenje.....	11
3.3 Struktura podataka unutar .npz datoteke.....	11
3.4 Broj instanci.....	11
3.5 Broj atributa.....	12
3.6 Broj klasa i njihovo tumačenje.....	12
3.7 Broj instanci po klasama.....	12
3.8 Količina podataka (MB).....	12
3.9 Vizuelna analiza i inspekcija podataka.....	12
3.10 Metode pretprocesiranja podataka.....	13
3.11 Identifikacija rizika.....	15
4 Odabir, formiranje, treniranje i testiranje modela.....	16
4.1 Odabir metode rješavanja problema.....	16
4.2 Izbor i opis korištenih tehnologija.....	16
4.3 Priprema formata podataka za odabrani model.....	17
4.4 Detaljan opis procesa treniranja modela.....	17
4.4.1 Definicija i modifikacija modela ResNet50.....	17
4.4.2 Funkcija gubitka, optimizator i raspoređivač stope učenja.....	17

4.4.3 Petlja treniranja.....	18
4.5 Evaluacija modela, rezultati i diskusija.....	18
4.5.1 Analiza konačnih performansi na testnom skupu.....	18
4.5.2 Analiza procesa učenja.....	20
4.5.3 Odgovor na identifikovane rizike.....	20
5 Zaključak.....	22
5.1 Osvrt na postignute rezultate.....	22
5.2 Poređenje sa stanjem u oblasti.....	22
5.3 Potencijalna unapređenja.....	22
6 Reference.....	24

1 Uvod i opis problema

Rak dojke je najčešći oblik karcinoma dijagnosticiran kod žena. Rano otkrivanje raka dojke ima ključnu ulogu u spašavanju života i smanjenju troškova liječenja. Klasične metode poput samopregleda, kliničkog pregleda koji vrši doktor i mamografije mogu dovesti do pogrešnih dijagnoza i lažno pozitivnih ili lažno negativnih nalaza. Zbog toga, javila se potreba za automatizovanim sistemima koji koriste metode vještačke inteligencije i dubokog učenja za analizu i klasifikaciju medicinskih slika.

Razvoj vještačke inteligencije, posebno u oblasti mašinskog i dubokog učenja, otvorio je brojne mogućnosti za unapređenje medicinske dijagnostike. Posebno se ističu konvolucione neuronske mreže (CNN), koje su pokazale veoma dobru sposobnost analize i klasifikacije vizuelnih podataka, uključujući i medicinske slike. Takvi sistemi ne treba da zamijene medicinsko osoblje, već da im služe kao dodatna podrška pri donošenju tačnih dijagnoza, naročito u pogledu ranog otkrivanja malignih promjena koje nisu tako lako uočljive ljudskim okom.

Cilj ovog rada je istražiti mogućnosti upotrebe jednog od najznačajnijih modela dubokog učenja, ResNet50, za klasifikaciju ultrazvučnih slika iz BreastMNIST skupa podataka, koji obuhvata ultrazvučne slike dojke klasifikovane kao benigne ili maligne.

1.1 Osnovni pojmovi

1.1.1 Osnovni tehnički pojmovi

Prema [4], definisani su sljedeći pojmovi:

Klasifikacija medicinskih slika - predstavlja proces u kojem se na osnovu vizuelnih karakteristika slike donosi odluka o njenoj pripadnosti unaprijed definisanoj dijagnostičkoj kategoriji.

Dataset - skup unaprijed pripremljenih i često stručno označenih podataka koje model koristi za učenje.

Labela – unaprijed dodijeljena oznaka (kategorija) koja definiše kojoj klasi određeni uzorak podataka (npr. slika) pripada. Koristi se kako bi model naučio da pravi ispravne predikcije.

Konvolucijske neuronske mreže (CNN – Convolutional Neural Networks) - posebna vrsta neuronskih mreža koje sadrže konvolucijske slojeve.

ResNet (Residual Networks) - jedna od najuticajnijih CNN arhitektura, koja uvodi tzv. rezidualne veze između slojeva.

BreastMNIST - dataset koji sadrži ultrazvučne slike dojke klasifikovane u dvije kategorije: benigno i maligno.

Pretprocesiranje podataka - skup tehnika koje pripremaju podatke za treniranje modela u cilju poboljšanja tačnosti.

Transfer učenje - metoda u mašinskom učenju gdje se model razvijen za jedan zadatak ponovo koristi kao polazna tačka za model na drugom, često srodnom zadatku.

Fino podešavanje (Fine-tuning) - proces adaptacije predtreniranog modela na novi, specifičan skup podataka.

Augmentacija podataka - tehnika koja se koristi za vještačko povećanje veličine i raznovrsnosti trening skupa podataka primjenom različitih transformacija na postojeće slike.

Funkcija gubitka (Loss Function) - matematička funkcija koja kvantifikuje razliku između predviđanja modela i stvarnih vrijednosti (labela). Cilj procesa treniranja je minimizacija ove funkcije.

Focal Loss - specijalizovana funkcija gubitka dizajnirana za rješavanje problema disbalansa klasa.

Optimizer - algoritam koji prilagođava težine neuronske mreže tokom procesa treniranja kako bi se minimizovala funkcija gubitka.

Stopa učenja - hiperparametar koji kontroliše veličinu koraka koje optimizator pravi prilikom ažuriranja težina modela.

Batch - podskup trening podataka koji se istovremeno propušta kroz model tokom jedne iteracije treniranja.

Epoha - jedan kompletan prolaz kroz cjelokupan trening skup podataka.

Pretreniranost (Overfitting) - fenomen gdje model postaje previše dobro prilagođen specifičnostima trening podataka.

ROC kriva (Receiver Operating Characteristic Curve) - grafički prikaz dijagnostičke sposobnosti binarnog klasifikatora pri različitim pragovima odluke.

AUC (Area Under the Curve) - površina ispod ROC krive. Predstavlja mjeru ukupne sposobnosti modela da razlikuje između pozitivne i negativne klase.

Matrica konfuzije - tabela koja prikazuje broj istinski pozitivnih (TP), istinski negativnih (TN), lažno pozitivnih (FP) i lažno negativnih (FN) predikcija.

1.1.1 Osnovni medicinski pojmovi

Prema [5], definisani su sljedeći pojmovi:

Benigna promjena - nije kancerogena, obično raste sporo i ne širi se na druge dijelove tijela.

Maligna promjena - kancerogena, može se proširiti na okolno tkivo ili druge dijelove tijela.

Biopsija - medicinski postupak u kojem se uzima uzorak tkiva dojke kako bi se laboratorijski analiziralo da li je promjena benigna ili maligna.

Ultrazvuk dojke - dijagnostička procedura koja koristi zvučne talase visoke frekvencije za stvaranje slika unutrašnjosti dojke.

Lažno pozitivan nalaz - rezultat testa koji pogrešno ukazuje na prisustvo bolesti kada je osoba zapravo zdrava.

Lažno negativan nalaz - rezultat testa koji pogrešno ukazuje na odsustvo bolesti kada je osoba zapravo bolesna.

1.2 Koristi rješavanja problema u kliničkoj praksi

Razvoj i primjena sistema za automatsku klasifikaciju mamografskih i ultrazvučnih slika donosi višestruke koristi, kako za pacijente, tako i za cjelokupni zdravstveni sistem. Jedna od

najvažnijih prednosti jeste rana detekcija karcinoma, budući da automatski sistemi omogućavaju pravovremeno prepoznavanje sumnjivih promjena koje bi u protivnom mogle ostati neprimijećene. Osim toga, ovi sistemi služe kao podrška radiolozima, djelujući kao drugo mišljenje i time smanjujući kognitivno opterećenje i rizik od dijagnostičkih grešaka.

Automatizacija dijagnostike također doprinosi većoj skalabilnosti i pristupačnosti, što znači da se visokokvalitetna analiza može pružiti i u sredinama koje nemaju dovoljno specijalizovanog medicinskog osoblja. Uz to, automatski sistemi doprinose standardizaciji dijagnostičkih odluka, smanjujući varijabilnost među različitim radiolozima i osiguravajući dosljednije rezultate. Konačno, ovakvi sistemi omogućavaju uštedu vremena i resursa, jer preuzimaju rutinske zadatke i time omogućavaju radiolozima da se fokusiraju na složenije i zahtjevnije slučajeve.

1.3 Pregled postojećih *dataset*-ova

Neki od postojećih *dataset*-ova navedeni su u sljedećoj tabeli:

Naziv	Opis	Format podataka	Link za preuzimanje
BreastMNIST	Sadrži 780 slika, postoji više verzija u zavisnosti od dimenzija slike. Sadrži 2 klase.	.npz	https://zenodo.org/records/10519652
BUSI	Sadrži 780 slika dimenzija 500x500 piksela. Sadrži 3 klase.	.png	https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset
CBIS-DDSM	Sadrži 2620 mamografskih snimaka. Sadrži 3 klase.	.png, .xml	https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

2 Pregled stanja u oblasti

Mamografija se i dalje smatra najučinkovitijom metodom za rano otkrivanje raka dojke, ali ručna analiza mamograma zahtijeva visok nivo stručnosti i podložna je ljudskim greškama, te je otežavaju visoke stope lažno pozitivnih i lažno negativnih rezultata. U tom kontekstu, računarska analiza mamografskih, kao i ultrazvučnih snimaka postaje sve značajnija, posebno uz napredak metoda dubokog učenja koje omogućavaju automatsko prepoznavanje i klasifikaciju lezija. Duboko učenje se pokazalo kao izuzetno efikasno u rješavanju problema klasifikacije slika i detekcije objekata, a u medicinskoj oblasti doprinosi razvoju sistema koji mogu služiti kao podrška radiolozima prilikom donošenja dijagnoza. Današnja istraživanja su usmjerena na razvoj pouzdanih i preciznih modela zasnovanih na dubokim konvolucijskim neuronskim mrežama, s ciljem povećanja tačnosti i brzine u dijagnostici raka dojke.

Prema istraživanju koje je objavljeno u časopisu *Nature* u januaru 2020. godine, te koje je rađeno nad velikim reprezentativnim skupom podataka iz UK i velikim obogaćenim skupom podataka iz SAD, pokazano je smanjenje pojave lažno pozitivnih rezultata za 5,7% (SAD) i 1,2% (UK), te lažno negativnih rezultata za 9,4% (SAD) i 2,7% (UK) korištenjem AI sistema za detekciju raka dojke [1].

Profesor Alexander Katalinic, koautor studije sa Univerziteta u Lubecku u Njemačkoj, izjavio je da je moguće poboljšati stopu detekcije raka dojke bez povećanja rizika za žene koje se izlažu *screening*-u. Katalinic i njegove kolege su analizirali podatke od 461.818 žena starosne dobi između 50 i 69 godina, koje su prošle *screening* raka dojke između jula 2021. i februara 2023. Godine. Sve žene su imale snimke koje su nezavisno pregledala dva radiologa. Međutim, kod 260.739 žena, najmanje jedan radiolog je koristio AI alat kao podršku pri analizi. Na kraju je zaključeno da je otkriven jedan dodatni slučaj raka na svakih 1.000 pregledanih žena kada je korištena vještačka inteligencija [2].

Prema istraživanju koje je objavljeno u časopisu *Radiology* u martu 2024. godine, istraživači su razvili novi model koji je u stanju da prepozna rizik od raka dojke na osnovu mamograma, te da predvidi putanju njegovog razvoja u periodu od pet godina. Mirai, savremeni algoritam zasnovan na dubokom učenju, pokazao je visoku sposobnost u predviđanju raka dojke, ali zbog

ograničenog razumijevanja njegovog načina zaključivanja postoji rizik da se radiolozi previše oslone na njega, što može dovesti do pogrešnih dijagnoza. „Mirai je crna kutija — veoma velika i složena neuronska mreža, slična po strukturi ChatGPT-u — i niko nije znao kako donosi svoje odluke“, rekao je glavni autor ove studije, Jon Donnelly, diplomirani inženjer i doktorand na Odsjeku za računarstvo Univerziteta Duke u Durhamu, Sjeverna Karolina [3].

3 Izbor, analiza i preprocesiranje *dataset*-a

3.1 Izvor *dataset*-a

Za potrebe ovog projekta izabran je BreastMNIST *dataset*. Ovaj skup podataka je dio šire i javno dostupne MedMNIST v2 kolekcije. Ovaj *dataset* je dostupan u različitim verzijama, sa veličinama slika od 28x28 piksela, 64x64 piksela i 224x224 piksela. Mi ćemo raditi sa skupom podataka koji sadrži slike veličine 224x224 piksela.

3.2 Format u kojem je dostupan i način za preuzimanje i korištenje

Skup podataka BreastMNIST je dostupan u .npz formatu, što ga čini pogodnim za obradu u Python okruženju. Preuzet je sa zvanične Zenodo arhive MedMNIST v2 projekta.

U ovom projektu, .npz datoteka pod nazivom breastmnist_224.npz je prethodno preuzeta i smještena na Google Drive. Učitavanje podataka izvršeno je pomoću numpy.load() funkcije i odgovarajuće putanje do .npz *file*-a.

3.3 Struktura podataka unutar .npz datoteke

Nakon učitavanja .npz datoteke, analiza sadržaja pokazuje da su podaci organizovani u logičke cjeline, identifikovane sljedećim ključevima: *train_images*, *train_labels*, *val_images*, *val_labels*, *test_images*, *test_labels*. Ovi ključevi omogućavaju direktno izdvajanje slika i njihovih pripadajućih labela za trening, validacioni i testni skup.

3.4 Broj instanci

Ukupan broj instanci (slika) u BreastMNIST skupu podataka iznosi 780. Skup podataka je unaprijed podijeljen na tri skupa, sa sljedećim brojem instanci:

- Trening skup: 546 instanci
- Validacioni skup: 78 instanci
- Testni skup: 156 instanci.

3.5 Broj atributa

Pošto je BreastMNIST slikovni skup, atributi nisu eksplicitne kolone kao u CSV *file*-u, odnosno tabelarnom prikazu podataka. Svaka vrijednost piksela se može posmatrati kao jedan atribut, pa je prema tome ukupan broj atributa po slici: $224 * 224 = 50.176$ atributa (piksela).

3.6 Broj klasa i njihovo tumačenje

Na osnovu zvanične dokumentacije MedMNIST skupa, što je potvrđeno i programskom analizom, utvrđeno je da postoje dvije klase:

- Klasa 0: Benigno,
- Klasa 1: Maligno.

3.7 Broj instanci po klasama

Distribucija klasa unutar BreastMNIST skupa podataka jasno pokazuje izražen disbalans između benignih i malignih slučajeva, što je vidljivo u sva tri podskupa.

U trening skupu, od ukupno 546 instanci, samo 147 uzoraka pripada klasi "benigno", dok čak 399 uzoraka pripada klasi "maligno", što znači da maligne promjene čine skoro tri četvrtine podataka za obuku.

U validacionom skupu, situacija je slična: od ukupno 78 instanci, svega 21 uzorak je benigni, dok su 57 uzoraka maligni.

Testni skup prati isti obrazac, gdje od ukupno 156 uzoraka, 42 su benigni, a 114 maligni.

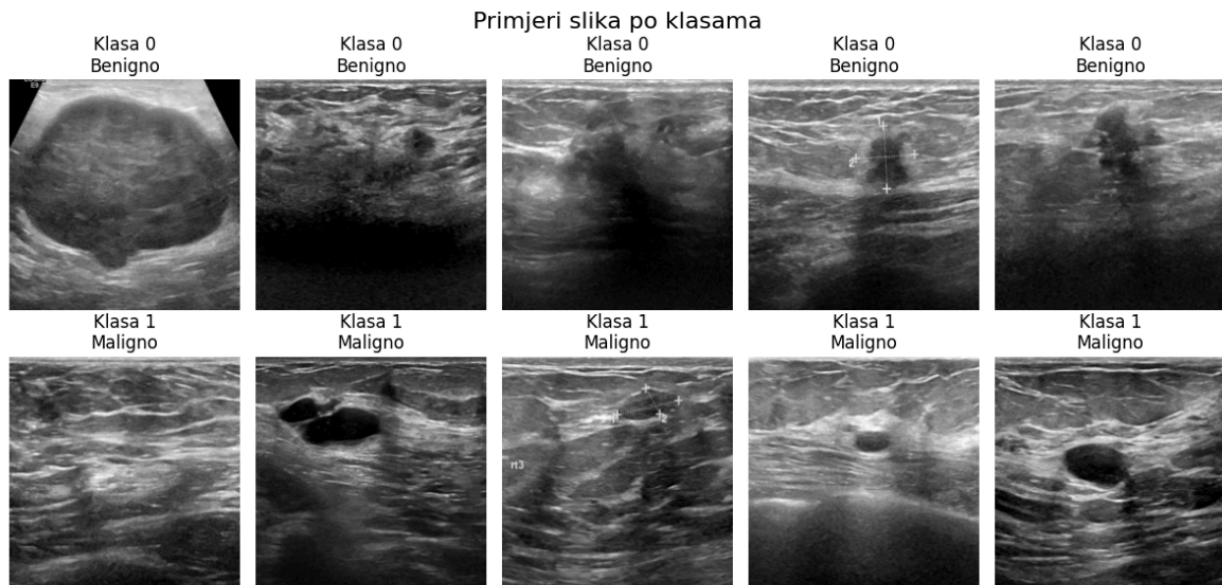
3.8 Količina podataka (MB)

Ukupna veličina slikovnih podataka (bez labela) u skupu podataka, izračunata na osnovu nbytes atributa NumPy nizova za trening, validacione i testne slike, iznosi približno 37.32 MB.

3.9 Vizuelna analiza i inspekcija podataka

Da bismo stekli bolji uvid u karakteristike slika sa kojima ćemo raditi i kako bismo potvrdili ispravnost učitanih podataka, izvršili smo vizuelnu inspekciju reprezentativnih uzoraka iz trening skupa. Korištenjem matplotlib.pyplot biblioteke u Python-u, prikazali smo nasumično odabrane primjere ultrazvučnih snimaka za svaku od definisanih klasa.

Prikaz slika po klasama:



3.10 Metode pretprocesiranja podataka

Uspjeh bilo kojeg projekta zavisi od kvaliteta ulaznih podataka koje dajemo algoritmima. Budući da razni ljudi, poslovni procesi i aplikacije često proizvode, obrađuju i skladište podatke iz stvarnog svijeta, ti podaci neminovno postaju haotični. Algoritmi ne mogu obraditi nepotpune podatke ili podatke koji imaju određeni šum jer obično nisu dizajnirani da upravljaju nedostajecim vrijednostima. Zbog svega toga, pretprocesiranje podataka je od izuzetne važnosti jer osigurava da će izlazni podaci biti pouzdani i precizni. Podaci sa kojima mi radimo na ovom projektu su slike, pa ćemo se fokusirati na metode pretprocesiranja slika.

Pretprocesiranje slike predstavlja unapređenje slikovnih podataka u cilju smanjenja neželjenih izobličenja ili poboljšanja određenih karakteristika slike koje su važne za daljnju obradu.

Pretprocesiranje se najčešće koristi za ispravljanje radiometrijskih ili geometrijskih grešaka, poboljšanje slike, njenu rekonstrukciju i kompresiju slikovnih podataka.

U okviru ovog projekta, faza pretprocesiranja ima za cilj transformaciju podataka u format optimalan za PyTorch framework i odabranu ResNet50 arhitekturu, te implementaciju strategija za poboljšanje kvaliteta učenja.

Prvi korak u obradi samih slikovnih podataka, nakon njihovog učitavanja, jeste normalizacija vrijednosti piksela. Originalne slike imaju vrijednosti piksela u opsegu od 0 do 255. Dijeljenjem sa 255.0 ove vrijednosti se skaliraju na standardni opseg [0.0, 1.0]. Istovremeno, tip podataka se konvertuje iz uint8 u float32.

Dalje, za efikasnu integraciju sa PyTorch DataLoader-om i primjenu transformacija, definiše se prilagođena klasa *BreastMNISTDataset* koja nasljeđuje *torch.utils.data.Dataset*. Transformacije slika imaju dvostruku ulogu: pripremaju podatke za ulaz u model i, što je još važnije, vještački povećavaju raznovrsnost trening skupa kroz augmentaciju. Ovaj korak je od presudnog značaja za uspjeh projekta, jer direktno utiče na sposobnost modela da generalizuje znanje i smanjuje rizik od *overfitting*-a.

Tehnike poput *RandomHorizontalFlip* (nasumično horizontalno okretanje), *RandomRotation* (rotacija), *RandomResizedCrop* (isijecanje i promjena veličine) i *ColorJitter* (varijacije u osvjetljenju/kontrastu) osiguravaju da model na svakoj epohi treniranja vidi blago izmijenjene verzije istih slika. Ovo ga primorava da uči suštinske, strukturne karakteristike lezija, umjesto da pamti specifične piksele i šum u trening podacima.

Dalje, primjenjujemo normalizaciju sa zadatim vrijednostima za srednju vrijednost (*mean* = 0.5) i standardnu devijaciju (*std* = 0.5). Ovim korakom se vrijednosti piksela iz opsega [0.0, 1.0] transformišu u opseg [-1.0, 1.0]. Standardizacija ulaznih podataka doprinosi stabilnijem i bržem procesu treniranja, jer osigurava da su ulazne vrijednosti centrirane oko nule.

Bitno je spomenuti da se na validacionom i testnom skupu ne primjenjuje augmentacija. Ovi skupovi služe za objektivnu i ponovljivu procjenu performansi modela. Na njih primjenjujemo samo neophodne transformacije (*ToTensor* i *Normalize*) kako bi podaci bili u istom formatu kao i trening podaci.

Finalni korak pripreme podataka je kreiranje *DataLoader* objekata, koji će modelu efikasno isporučivati podatke u *batch*-evima, uz implementaciju strategije za balansiranje klasa. Dakle, upotrijebljen je *WeightedRandomSampler* za trening skup kako bi se ublažio efekat disbalansa klasa, osiguravajući da model tokom treniranja ne zanemari manjinsku, benignu klasu.

3.11 Identifikacija rizika

Tokom analize BreastMNIST *dataseta* identifikovano je nekoliko ključnih rizika koji mogu uticati na performanse modela:

- Vizuelna sličnost između klasa: Slike benignih i malignih promjena često imaju sličnu teksturu i intenzitet, što otežava tačnu klasifikaciju.
- Disbalans klasa: Dataset sadrži znatno više malignih slučajeva, što može dovesti do toga da model nauči da jednostavno uvijek predviđa većinsku klasu kako bi postigao visoku ukupnu tačnost.
- Varijabilnost unutar iste klase: Snimci unutar iste dijagnostičke klase (benigno ili maligno) mogu se značajno razlikovati po pitanju kontrasta, jasnoće prikaza anatomskih struktura i pozicije lezije.
- Rizik od *overfitting*-a: S obzirom na relativno mali broj jedinstvenih trening uzoraka (546) u poređenju sa kapacitetom ResNet50 arhitekture (brojem parametara koje može naučiti), postoji opasnost da model previše dobro nauči specifičnosti trening skupa, uključujući i šum, umjesto da generalizuje na nove podatke.

4 Odabir, formiranje, treniranje i testiranje modela

4.1 Odabir metode rješavanja problema

Za klasifikaciju slika, metoda dubokog učenja sa konvolucionim neuronskim mrežama predstavlja standardni pristup. U ovom projektu, primijenjena je tehnika transfer učenja sa ResNet50 arhitekturom. ResNet50 je odabran jer može učiti složenije obrasce i detalje u poređenju sa jednostavnijim arhitekturama, što je posebno korisno za prepoznavanje suptilnih razlika na medicinskim slikama.

Kako bi se efikasno riješio problem disbalansa klasa identifikovan u Fazi 3, primijenjen je dvostruki pristup:

- *WeightedRandomSampler*: Na nivou podataka, osigurava da model u svakoj epohi vidi balansiran broj primjera iz obje klase.
- *Focal Loss*: Na nivou modela, ova funkcija gubitka primorava model da se fokusira na teže, često pogrešno klasifikovane primjere, koji obično pripadaju manjinskoj klasi.

4.2 Izbor i opis korištenih tehnologija

Za realizaciju ovog projekta korišten je sljedeći skup tehnologija i alata:

- Python - odabran kao osnovni programski jezik zbog svoje svestranosti, čitljivosti i izuzetno bogatog izbora biblioteka specijalizovanih za nauku o podacima, mašinsko učenje i numeričko računanje.
- PyTorch (torch, torch.nn, torch.optim) - korišten kao primarni framework za duboko učenje.
- Torchvision (torchvision.models, torchvision.transforms)- služi kao ključna ekstenzija PyTorch-a, specifično namijenjena zadacima iz oblasti računarske vizije.
- Google Colaboratory (Colab) - cjelokupan razvoj, uključujući pisanje koda, eksperimentisanje i dugotrajne procese treniranja modela, izvršen je unutar Google Colaboratory okruženja. Njegova glavna prednost je besplatan pristup GPU resursima

(poput NVIDIA Tesla T4), što je esencijalno za značajno ubrzanje treniranja dubokih neuronskih mreža.

4.3 Priprema formata podataka za odabrani model

Format podataka je u potpunosti definisan i pripremljen u Fazi 3. Rezultat te faze su PyTorch DataLoader objekti koji služe kao most između sirovih podataka i modela, osiguravajući efikasno i optimalno isporučivanje podataka tokom treniranja i evaluacije. DataLoader instance su dostupne i konfigurisane tako da isporučuju podatke u *batch*-ovima od 32 slike. Slike koje ulaze u model su PyTorch tenzori oblika (1, 224, 224), što podrazumijeva jedan kanal za grayscale, visina 224 i širina 224 sa vrijednostima piksela normalizovanim na opseg [-1.0, 1.0]. Adresiran je i problem disbalansa klasa, što je detaljno opisano u prethodnoj fazi.

Ova sveobuhvatna priprema osigurava da su podaci optimalno formatirani za ResNet50 arhitekturu i napredni proces treniranja koji slijedi.

4.4 Detaljan opis procesa treniranja modela

4.4.1 Definicija i modifikacija modela ResNet50

Model se inicijalizuje kao predtrenirana ResNet50 arhitektura. Prvi konvolucionni sloj (conv1) je zamijenjen kako bi prihvatio *grayscale* slike, a izlazni potpuno povezani sloj (fc) se modifikuje da ima jedan izlazni neuron za binarnu klasifikaciju. Primijenjena je strategija finog podešavanja gdje su težine ranijih slojeva zamrznute, dok se težine slojeva layer2, layer3, layer4 i fc sloja treniraju kako bi se model adaptirao na specifični medicinski dataset.

4.4.2 Funkcija gubitka, optimizator i raspoređivač stope učenja

Kao funkcija gubitka koristi se *FocalLoss* sa parametrima $\alpha=0.25$ i $\gamma=2$, kako bi se efikasno adresirao disbalans klasa. Za optimizaciju se koristi Adam algoritam sa stopom učenja $3e-4$ i *weight_decay* od $1e-4$ kao oblik regularizacije. Stopa učenja se dinamički prilagođava pomoću *ReduceLROnPlateau* raspoređivača.

4.4.3 Petlja treniranja

Model se trenira u maksimalno 30 epoha. U svakoj epohi, vrši se treniranje na trening skupu i evaluacija na validacionom skupu. ROC AUC na validacionom skupu se koristi kao glavna metrika za praćenje napretka i za mehanizam ranog zaustavljanja (patience = 10). Stanje modela sa najboljim validacionim AUC-om se čuva.

Analiza izlaza procesa treniranja pokazuje da je model uspješno učio tokom svih 30 definisanih epoha. Iako mehanizam ranog zaustavljanja nije aktiviran (jer se validacioni AUC nije pogoršao za 10 uzastopnih epoha), ključno je što je sačuvano stanje modela iz epohe sa najvišom AUC vrijednošću na validacionom skupu (AUC=0.9599 u 21. epohi). Upravo se ovaj, optimalni model koristi za finalnu evaluaciju, a ne model iz posljednje epohe.

4.5 Evaluacija modela, rezultati i diskusija

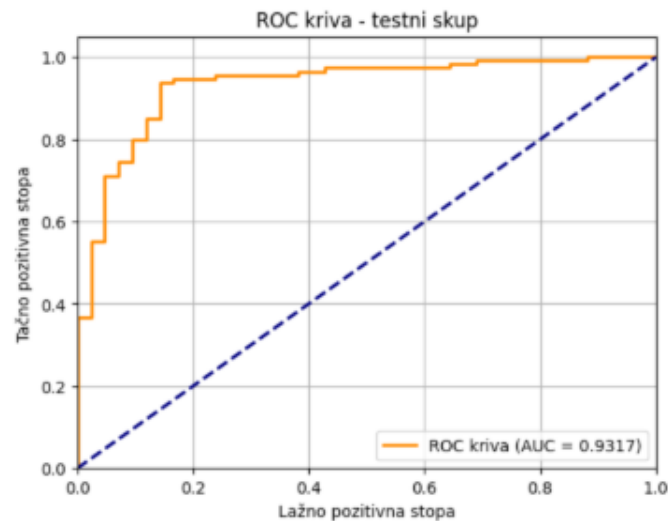
Nakon završenog procesa treniranja, model sa najboljim performansama na validacionom skupu (onaj sačuvan kao `best_resnet50_finetuned.pth`) je podvrgnut evaluaciji na testnom skupu. Ovaj proces omogućava objektivnu procjenu njegove konačne sposobnosti generalizacije na potpuno neviđenim podacima. Analiza performansi obuhvata kako konačne metrike na testnom skupu, tako i uvid u dinamiku učenja kroz epohe.

4.5.1 Analiza konačnih performansi na testnom skupu

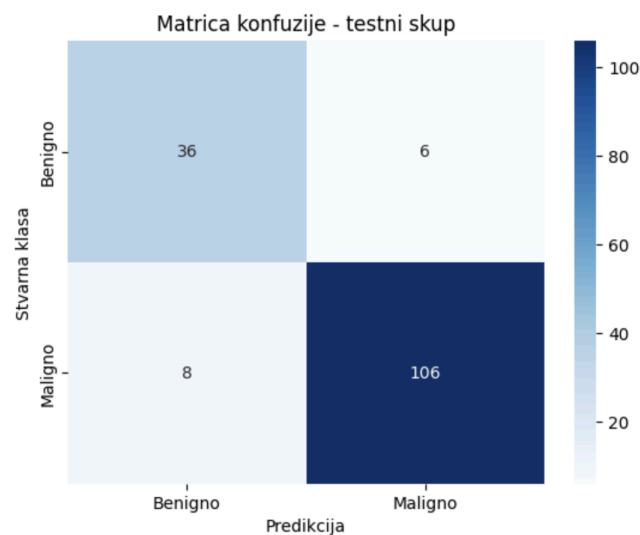
Ključan korak u evaluaciji bio je određivanje optimalnog klasifikacionog praga analizom ROC krive na testnom skupu. Prag koji maksimizira razliku između stope tačno pozitivnih i lažno pozitivnih nalaza postavljen je na 0.4095. Korištenjem ovog praga, dobijene su sljedeće konačne metrike performansi:

- Ukupna tačnost (*accuracy*) od 91.03% znači da je sistem ispravno klasifikovao više od 91% instanci svih uzoraka u testnom skupu, što je u oblasti klasifikacije medicinskih slika izuzetno dobar rezultat.
- *F1-score* metrika za klasu “maligno” iznosi 0.9381, što ukazuje na vrlo dobar balans između preciznosti i odziva za tu klasu.

- ROC AUC vrijednost iznosi 0.9317. Ova metrika se vizuelno predstavlja ROC krivom koja se izdiže ka gornjem lijevom uglu, te kvantitativno potvrđuje izvanrednu sposobnost modela da razlikuje između benigne i maligne klase, daleko nadmašujući nasumično pogađanje.

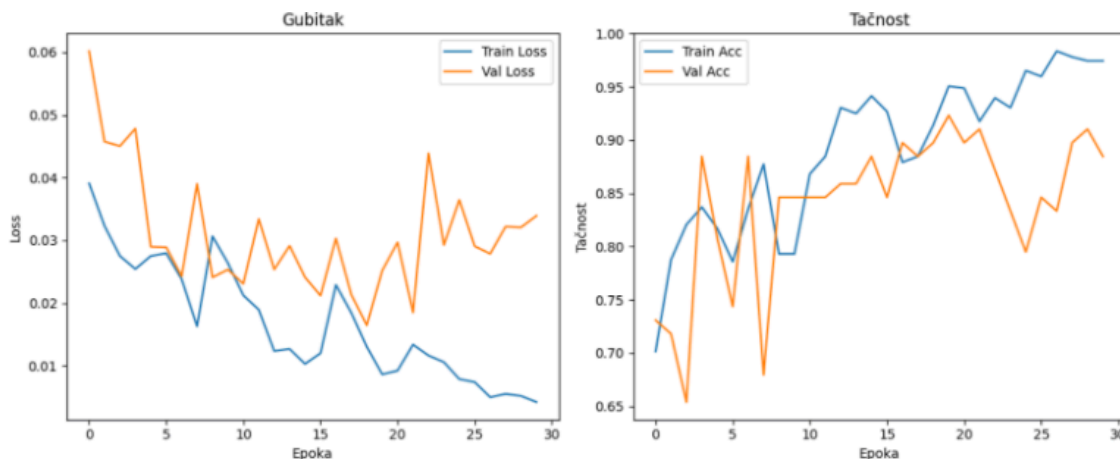


- Izvještaj o klasifikaciji dodatno pojašnjava performanse po klasama:
 - a. Za benignu klasu preciznost iznosi 0.82, odziv 0.86 a F1-score 0.84
 - b. Za malignu klasu preciznost iznosi 0.95, odziv 0.93, a F1-score 0.94
- Matrica konfuzije pokazuje da je model tačno klasifikovao 106 malignih i 36 benignih slučajeva, dok je 8 malignih pogrešno klasifikovao kao benigne i 6 benignih kao maligne.



4.5.2 Analiza procesa učenja

Analiza krivih gubitka (Loss) i tačnosti (Accuracy) tokom epoha, prikazanih na odgovarajućim grafovima, pruža uvid u dinamiku procesa učenja i sposobnost generalizacije razvijenog modela.



Performanse na trening skupu (prikazane plavom linijom) pokazuju uspješan proces optimizacije modela. Kontinuirani pad funkcije gubitka (*Train Loss*) uz istovremeni rast tačnosti (*Train Acc*) potvrđuje da je model uspješno učio i prilagođavao svoje naučene parametre na osnovu trening podataka.

Sa druge strane, metrike na validacionom skupu (prikazane narandžastom linijom) pokazuju izraženu nestabilnost. Ove oscilacije nisu pokazatelj neuspješnog treniranja, već očekivana posljedica rada sa malim validacionim skupom (78 uzoraka). Ključno je uočiti da, uprkos ovim oscilacijama, validacioni gubitak ne raste kontinuirano. Ovo pokazuje da su primijenjene tehnike regularizacije, prvenstveno *weight_decay* i *augmentacija* podataka, bile efikasne u smanjenju rizika od *overfittinga*.

4.5.3 Odgovor na identifikovane rizike

Sveukupna analiza rezultata potvrđuje da je odabrana metodologija uspješno odgovorila na ključne rizike identifikovane u Fazi 3. Kombinovani pristup *WeightedRandomSampler*-a i *Focal Loss* funkcije pokazao se izvrsnim kod problema disbalansa klasa, omogućivši visok odziv (86%) za manjinsku, benignu klasu i spriječivši pristrasnost modela.

Korištenje duboke ResNet50 arhitekture omogućilo je modelu da nauči suptilne, ali ključne diskriminativne karakteristike.

Kao što je i ranije spomenuto, tehnike regularizacije i augmentacije su efikasno držale rizik od *overfitting*-a pod kontrolom.

5 Zaključak

Ovaj projekat je imao za cilj razvoj i evaluaciju naprednog sistema za automatsku klasifikaciju ultrazvučnih slika dojke, koristeći duboku konvolucionu neuronsku mrežu. Kroz sistematičan pristup, od analize podataka do implementacije modela, uspješno je demonstrirana visoka efikasnost vještačke inteligencije u rješavanju ovog problema.

5.1 Osvrt na postignute rezultate

Implementirani model, zasnovan na ResNet50 arhitekturi uz tehniku transfer učenja, postigao je veoma dobre performanse na testnom skupu podataka. Ključni pokazatelji uspjeha su visoka ukupna tačnost od 91.03%, te jako dobar balans između preciznosti i odziva.

Posebno je značajan uspjeh u rješavanju problema disbalansa klasa. Kombinovana strategija WeightedRandomSampler-a (na nivou podataka) i Focal Loss funkcije (na nivou modela) rezultirala je visokim odzivom (recall) od 86% za benignu klasu. Ovo je ključno, jer model efikasno izbjegava da bude pristrasan prema većinskoj klasi, što je čest izazov u medicinskim datasetovima.

5.2 Poređenje sa stanjem u oblasti

U Fazi 2 analizirani su radovi koji su pokazali da AI sistemi mogu značajno smanjiti stope lažno pozitivnih i lažno negativnih nalaza. Postignuti rezultati našeg modela su visoko konkurentni i u potpunosti u skladu sa modernim AI rješenjima opisanim u literaturi. Vrijednost ROC AUC metrike od 0.9317 ukazuje na visoku efikasnost razvijenog pristupa u kontekstu medicinske dijagnostike, gdje se vrijednosti iznad 0.90 generalno smatraju vrlo dobrim rezultatima.

Model je pokazao sposobnost učenja suptilnih, klinički relevantnih karakteristika, što ga čini uporedivim sa sistemima koji se danas istražuju i primjenjuju.

5.3 Potencijalna unapređenja

Iako su postignuti rezultati vrlo dobri, postoji nekoliko načina na koje bi se ovaj rad mogao unaprijediti. Naime, iako je broj lažno negativnih predikcija bio mali (8 slučajeva), buduća istraživanja bi se mogla usmjeriti na njegovo dodatno smanjenje – na primjer, prilagođavanjem funkcije gubitka. Također, većina hiperparametara je odabrana prema ustaljenim pravilima, pa bi sistematska optimizacija mogla dodatno poboljšati rezultate. Osim toga, primjena naprednijih metoda za augmentaciju podataka moglo bi dodatno poboljšati generalizaciju modela. I na kraju, iako je model pokazao dobre rezultate na izdvojenom testnom skupu, za potpuniju provjeru njegove pouzdanosti bilo bi korisno testirati ga na većim i raznovrsnijim medicinskim *dataset*-ovima.

6 Reference

- [1] Scott Mayer McKinney, “Nature”, 2020. Dostupno:
<https://www.nature.com/articles/s41586-019-1799-6>
- [2] Nicola Davis, “The Guardian”, 2025. Dostupno:
<https://www.theguardian.com/society/2025/jan/07/more-breast-cancer-cases-found-when-ai-used-in-screenings-study-finds>
- [3] Jon Donnelly, “Radiological Society of North America RSNA”, 2024. Dostupno:
<https://www.rsna.org/news/2024/march/deep-learning-for-predicting-breast-cancer>
- [4] K. He, X. Zhang, S. Ren i J. Sun, “Deep Residual Learning for Image Recognition”, 2016
- [5] A. Becker et al., “Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study”, 2018