

EVALUACIJA KOMPAKTNIH MODELA DUBOKIH NEURONSKIH MREŽA ZA
KLASIFIKACIJU RENDGENSKIH SNIMAKA GRUDNOG KOŠA
EVALUATION OF COMPACT DEEP NEURAL NETWORK MODELS
FOR CHEST X-RAY IMAGE CLASSIFICATION

Vojislav Stevanović, Pavle Milošević

REZIME: Detekcija abnormalnosti na rendgenskim snimcima grudnog koša je izazovan zadatak prvenstveno usled vremenskih ograničenja i prevelike opterećenosti medicinskih stručnjaka. Neravnomerna raspodela i povremeno nepouzdanе labele dodatno otežavaju pouzdanu izgradnju automatizovanih sistema baziranih na neuronskim mrežama. U ovom radu istražene su tri kompaktne arhitekture neuronskih mreža: *MobileNetV4*, *EfficientFormerV2* i *EdgeNeXt* i upoređene su njihove performanse na ovom zadatku. Korišćenjem binarne klasifikacije i *focal* gubitka realizovana je adaptacija svakoj od abnormalnosti posebno, time omogućavajući drugačiji pristup od češće korišćene *multi-label* metode. Modeli su obučavani nad *NIH ChestX-ray14* skupom podataka uz trostruku stratifikovanu unakrsnu validaciju i minimalne augmentacije u korist očuvanja strukture slika. *EdgeNeXt* i *MobileNetV4* ostvaruju konkurentne performanse u odnosu na veće, tradicionalne mreže, sa prosečnim *ROC AUC* vrednostima od 79,59% i 77,73%, dok *EfficientFormerV2* postiže 70,95%. Rad potvrđuje potencijal efikasnih modela za primenu u kliničkim uslovima i otvara mogućnosti integracije automatskog generisanja izveštaja kao i objašnjivosti modela.

KLJUČNE REČI: kompjuterska vizija, klasifikacija slika, rendgenske slike grudnog koša, kompaktne neuronske mreže.

ABSTRACT: Detection of abnormalities in chest X-ray images is primarily challenging due to time constraints and the excessive workload of medical professionals. The uneven distribution and occasionally unreliable labels further complicate the reliable construction of automated systems based on neural networks. In this paper, three compact neural network architectures are investigated: *MobileNetV4*, *EfficientFormerV2*, and *EdgeNeXt*, and their performance is compared. Using binary classification and focal loss, adaptation to each abnormality is implemented separately, thus enabling a different approach than the more commonly used *multi-label* method. The models were trained on the *NIH ChestX-ray14* dataset with three-fold stratified cross-validation and minimal augmentation in favor of preserving image structure. *EdgeNeXt* and *MobileNetV4* achieve competitive performance compared to larger, traditional networks, with average *ROC AUC* values of 79.59% and 77.73%, while *EfficientFormerV2* achieves 70.95%. This confirms the potential of efficient models for application in clinical settings and opens up the possibility of integrating automatic report generation and model explanation.

KEY WORDS: computer vision, image classification, chest X-ray images, efficient neural networks.

1. UVOD

Detekcija abnormalnosti na radiografskim snimcima grudnog koša relevantan je istraživački i klinički problem usled činjenice da su rendgenski snimci pluća jedni od najčešće korišćenih u procesu dijagnostifikacije. Interpretacija snimaka ove vrste zahteva nezanemarivo znanje i iskustvo, a medicinsko osoblje često je preopterećeno brojem pregleda, kao i vremenski, pa to utiče na tačnost i doslednost dijagnoza (Kasalak et al., 2023). Razvoj algoritama dubokog učenja otvara mogućnost za automatizaciju analiza koje mogu služiti kao pomoćni alat radiolozima u dijagnostici različitih poremećaja, kao na primer zapaljenja pluća, srčanog uvećanja, nodula i sl. Osim što se razvitkom ovakve vrste sistema povećava efikasnost, oni mogu doprineti dostupnosti zdravstvene zaštite u okruženjima gde je ograničen broj specijalista (Topol, 2019). Ipak, zadatak je metodološki složen, uglavnom zbog nepouzdanosti oznaka u velikim skupovima podataka, kao i neravnomerne zastupljenosti patologija.

Prema skorijim analizama, zapaljenje i rak pluća, zajedno sa tuberkulozom će do 2030. predstavljati jedan od glavnih uzroka smrtnosti. Procena takođe govori o tome da će činiti oko jednu petinu smrti širom sveta (European Respiratory Society, 2023). Enorman teret ovog tipa pokazuje potrebu za širenjem dostupnosti dijagnostičkih alata. Dodatna činjenica koja ovo potvrđuje je da se svake godine obavi oko 3,5 milijardi rendgenskih snimaka, konstituirajući najveći deo dijagnostike koja se odvija

kroz slikovnu radiografiju (Mitchell, 2012). Uzevši sve ovo u obzir, razvoj pouzdanih i potencijalno objašnjivih modela i dalje je relevantna oblast istraživanja i implementacije.

Od 2017. pa do sada, došlo je do raznih pomaka u automatizovanoj analizi rendgenskih snimaka grudnog koša. Počelo je prvim javno dostupnim skupom podataka *NIH ChestX-ray14* i početnom analizom koja je koristila konvolucione neuronske mreže (engl. *Convolutional Neural Networks*, *CNNs*) mreže i transferno učenje (Wang et al., 2017). Dalje se problem razvio prikupljanjem dodatnih skupova kao što je *CheXpert* (Irvin et al., 2019), gde je unesena nova dimenzija „nesigurnih“ labela, što je otvorilo nove mogućnosti. U moderno vreme, studije se često oslanjaju na transformer ili hibridne *CNN*-transformer arhitekture. Ipak, nasuprot svim evolucijama, akcenat je retko na interpretabilnosti i korišćenju sistema u kliničkim kontekstima. Većina radova u literaturi koristi velike arhitekture koje su resursno zahtevne i nefleksibilne za ponovno obučavanje. Zajedno sa velikim arhitekturama dolazi i manjak diskusije o primeni modela u realnim slučajevima. Zbog toga ćemo u ovom radu istražiti odnose performansi i zauzetih resursa. Pristup u studijama je univerzalno *multi-label*, što je opravdano međuzavisnošću klasa, ali mi ćemo se voditi binarnim principom kako bismo se mogli svakoj abnormalnosti posebno posvetiti. Uzevši ovo u obzir, postoji jasna potreba za istraživanjem efikasnih modela i radu na interpretiranju rezultata kroz prizmu okruženja ograničenih resursa.

U ovom radu koristićemo tri kompaktne arhitekture neuronskih mreža, po tipu *CNN* i vizuelni transformer (engl. *vision transformer*, *ViT*), da bismo saznali da li su arhitekture evoluirale dovoljno da bi parirale prethodnim rešenjima. Istražićemo *MobileNetV4*, *EfficientFormerV2* i *EdgeNeXt* arhitekturu, koje će biti obučavane nad *NIH ChestX-ray14* skupom podataka, korišćenjem unakrsne validacije sa tri particije i podešavanjem funkcije gubitka po potrebi (Qin et al., 2024; Li et al., 2023; Maaz et al., 2022). Ove arhitekture izabrane su jer imaju mali broj parametara u odnosu na modele korišćene u literaturi, i zato što imaju dobru generalizaciju na skupovima podataka koji imaju ograničenu reprezentaciju za određene klase. Takođe ćemo se dohvatiti adaptacije svakoj od abnormalnosti individualno, da bismo videli da li je pristup binarne klasifikacije opravdan. Na ovaj način, istraživanjem različitih arhitektura efikasnih neuronskih mreža, daje se osnova evolucije korišćenja modela u uslovima ograničenih mogućnosti i realnim kliničkim kontekstima.

Rad će biti organizovan u četiri dodatne celine. Prvo će se preći teorijske osnove neuronskih mreža, kao i individualne karakteristike svake od korišćenih arhitektura. Zatim, biće dat pregled relevantne literature u kojoj će takođe biti prikazan pregled rezultata tih rešenja i kratak opis svakog od pristupa, a nakon toga doći će istraživanje skupa podataka i postavke problema. Završićemo poglavljem o eksperimentima i njihovim rezultatima, u kojem će biti reči o metodološkoj postavci eksperimenata, tehnikama prevazilaženja neuravnoteženosti klasa, postavci obučavanja modela, metrikama performansi i testiranju modela na odvojenom skupu podataka.

2. TEORIJSKE OSNOVE

U ovom poglavlju biće dato teorijsko objašnjenje za neuronske mreže koje su korišćene u problemu klasifikacije abnormalnosti na osnovu rendgenskih slika grudnog koša. Započeće se sa osnovnim objašnjenjem dubokih neuronskih mreža, kao i varijanti, pa će se dublje istražiti kompaktne neuronske mreže.

2.1. Duboke neuronske mreže, CNN i ViT mreže

Duboke neuronske mreže su modeli sa više uzastopnih slojeva koji omogućavaju učenje hijerarhijskih reprezentacija podataka (Bengio, 2009). Model tokom učenja postaje sposoban za aproksimaciju kompleksnih funkcija usled svoje dubine i velikog broja nelinearnih transformacija. Neuronske mreže prilikom analize slike omogućavaju ekstrakciju kompleksnih latentnih obrazaca na slici, direktno iz sirovih podataka, tj. piksela (Krizhevsky et al., 2012).

Konvolucione neuronske mreže specijalizovane su klasa neuronskih mreža koja je dizajnirana za rad sa prostorno strukturiranim podacima kao što su slike (LeCun et al., 2002; LeCun et al., 2015). Osnovna operacija je konvolucija koja putem lokalnih filtera omogućava ekstrakciju prostornih karakteristika znatno smanjenim brojem parametara u odnosu na mreže sa potpuno povezane slojeve. Na početnim slojevima, mreža uči ivice, dok se na kasnijim slojevima formiraju

apstraktnije forme kao što su anatomske obrasce i njihovi međusobni odnosi. Zbog efikasnosti i adaptibilnosti ovakvog tipa mreža, *CNN* modeli i dalje predstavljaju jednu od dominantnijih opcija u analizi medicinskih slika (Litjens et al., 2017).

Vizuelni transformer mreže prenose mehanizam pažnje iz oblasti prirodnog jezika na zadatke koji se tiču vizuelnih podataka (Dosovitskiy et al., 2020). Slika se deli na niz oblasti, koji se linearno projektuju u sekvencu vektora i onda obrađuju kroz slojeve multi-head self-attention mehanizma. U odnosu na *CNN* mreže koje uče lokalne obrasce putem konvolucija, *ViT* mreže modeluju globalne zavisnosti između svih delova slike u svakoj fazi obrade. Ova osobina može biti vrlo korisna tamo gde su relacije između udaljenih anatomske regiona klinički relevantne. Transformer mreže ipak zahtevaju veću količinu podataka i računarskih resursa, te nisu uvek očigledan izbor, iako na mnogim problemima nadmašuju performanse *CNN* modele.

Hibridne arhitekture kombinuju konvolucione slojeve sa mehanizmom pažnje kako bi se iskoristile prednosti obe ideje: i lokalne ekstrakcije karakteristika i globalnog modelovanja međuzavisnosti. Uobičajen je obrazac da se *CNN* koristi za prvobitno razdvajanje reprezentacija na nižim nivoima, dok se transformer slojevi nastupaju kasnije, radi integracije informacija na globalnom nivou. Razvoj efikasnih arhitektura koje omogućavaju implementaciju na mobilnim i *edge* uređajima je relevantan u medicinskim kontekstima (Amin & Hossain, 2020). Efikasne arhitekture takođe čine mogućom i širu dostupnost sistema veštačke inteligencije u realnim okruženjima, koja su neretko okarakterisana ograničenim resursima i infrastrukturom.

2.2. MobileNetV4

MobileNetV4 predstavlja četvrtu iteraciju u *MobileNet* porodici efikasnih konvolucionih neuronskih mreža. Evolucija je započela separabilnim konvolucijama u *MobileNetV1*, zatim sa obrnutim rezidualnim i linearnim *bottleneck* blokovima u *MobileNetV2* (Sandler et al., 2018) i *NAS* optimizacijom u *MobileNetV3* (Howard et al., 2019). Osnova četvrte edicije širi obrnutu *bottleneck* paradigmu na tzv. *Universal Inverted Bottleneck* (*UIB*) blok. Cilj dizajna ovog bloka je da bude adaptibilan prema različitim akceleratorima uređaja, bilo to centralna (engl. *central processing unit*, *CPU*), grafička (engl. *graphical processing unit*, *GPU*) ili neuronska procesorska jedinica (engl. *neural processing unit*, *NPU*). Prethodne iteracije bile su vezane za optimizaciju na osnovu broja operacija po sekundi (engl. *Floating-Point Operations Per Second*, *FLOPs*), dok četvrta pristupa svakom hardveru zasebno (engl. *hardware-aware*). Motivacija za nastankom nove iteracije u odnosu na *MobileNetV3* bila je jasna - *MobileNetV4* uvodi univerzalniji dizajn i time na različitim platformama omogućava bolji odnos tačnosti i latencije (Qin et al., 2024).

U ovoj arhitekturi kombinuju se efikasni konvolucionni slojevi sa komponentama koje su inspirisane mehanizmom pažnje (npr. *Mobile MQA*), ali bez potpunog prelaska na transformer dizajn. Duž razvoja mreže korišćena je ručna optimizacija u kombinaciji sa pretragom neuronske arhitekture (engl. *neural architecture search*, *NAS*), a takođe se vodilo računa i

o latenciji na različitim uređajima, i to je ulazilo u kriterijum optimizacije. Na ovaj se način otklonilo ograničenje prethodnih iteracija gde se optimizacija vršila samo na određenim tipovima procesora.

2.3. EfficientFormerV2

EfficientFormer se nadovezuje na ideju da se arhitekture bazirane na transformerima mogu porediti sa *CNN* modelima, uz određene arhitekturne adaptacije. U prvoj iteraciji ovog modela došlo je do pokušaja redukcije složenosti standardne *ViT* arhitekture (Dosovitskiy et al., 2021) i hijerarhijskih modela kao što je *Swin Transformer* (Liu et al., 2021). *EfficientFormerV2* zadržava hijerarhijsku strukturu koja je slična *CNN* mrežama, ali inkorporira mehanizam pažnje. Motivacija u odnosu na *EfficientFormerV1* ticala se redukcije i optimizacije blokova, kao i poboljšanje tačnosti bez povećavanja broja parametara, i samim tim, računске složenosti.

EfficientFormerV2 inovacija se svodi na optimizaciju blokova pažnje i višeslojnih perceptrona (engl. *multi-layer perceptron*, *MLP*) u odnosu na ograničenja mobilnih uređaja. Autori govore o tendenciji *ViT* arhitekture da imaju neefikasne obrasce pristupanja memoriji uređaja, time značajno degradirajući performanse u realnom vremenu. *EfficientFormerV2* se trudi da to ispravi, restrukturiranjem blokova radi smanjenja memorijskog *overhead*-a i postizanja latencije koja je slična *MobileNet* mrežama (Li et al., 2023).

Ovim je demonstrirano da arhitektura koja je bazirana na mehanizmima pažnje može dostići *CNN* modele u kontekstu mobilnih operacija, po pitanju latencije i tačnosti, i sve to, bez totalnog oslanjanja na konvolucije kao osnovne operacije unutar mreže.

2.4. EdgeNeXt

EdgeNeXt kombinuje konvolucione i transformer komponente, u ideji smanjenja računarske složenosti. Sa ciljem smanjenja kompleksnosti uobičajene *self-attention* operacije, koja je svojstvena *ViT* mrežama, pokušaj ove arhitekture se svodi na integrisanje mehanizma pažnje u efikasnu *CNN* osnovu (engl. *backbone*). Za razliku od čistih *ViT* modela, *EdgeNeXt* se drži konvolucionih slojeva u ranijim fazama obrade. Motivacija rada (Maaz et al., 2022) je bila slična kao za ostale mobilne transformere, a to je da se osmisli model koji će nadmašiti klasične *CNN* mreže po tačnosti, bez pogoršavanja na polju latencije.

Split Depth-wise Transpose Attention (SDTA) blok je u centru ovog modela. Deljenjem kanala u više grupa primenjuje efikasniju pažnju nad njima, zajedno u kombinaciji sa *depthwise* konvolucijama. Globalni kontekst koji je time zadržan značajno je manje računski složen u odnosu na standardne mehanizme *multi-head self-attention* slojeva. Slično kao u *CNN* mrežama, postoji hijerarhijska struktura, smanjenjem prostorne rezolucije duž mreže. Potvrđeno je da ova arhitektura postiže konkurentne performanse uz malu potrošnju resursa.

2.5. Poređenje arhitektura

U ovom radu biće predstavljene tri arhitekture, a svaka od navedenih arhitektura ima više različitih verzija, te je neophodno specificirati konkretne. Verzije arhitektura izabrane su sa pretpostavkom da se mogu izvršavati i na uređajima sa ekstremno ograničenim uslovima, te je nad njima imponovano ograničenje od 12 MB. U Tabeli 1. prikazane su konkretne verzije modela, kao i neke od njihovih karakteristika.

Tabela 1. Pregled modela i njihovih karakteristika

Model	Broj parametara (milioni)	Veličina u MB	GFLOPs	Top-1 tačnost na ImageNet skupu	Godina nas-tanka
<i>MobileNetV4 Conv Small E2400</i>	2,5	9	~ 0,2 – 0,4	74 - 75 %	2024
<i>EfficientFormer V2 S0</i>	3,2	12	~ 0,4	75,7 %	2023
<i>EdgeNeXt X Small</i>	2,1	8	~ 0,54	74 - 75 %	2022

Jasno primećujemo činjenicu da arhitekture ne poseduju veliki broj parametara, ali, nasuprot tome, postižu prilično konkurentnu tačnost na *ImageNet* testu. *EdgeNeXt* je po broju parametara najkompaktnija arhitektura, ali ne i po *GFLOPs* metrici, po kojoj je najefikasniji *MobileNetV4*. Najveću tačnost na *ImageNet* testu postiže *EfficientFormerV2*, mada su razlike po tom pitanju relativno male.

3. PREGLED LITERATURE

Na početku realizacije prvih automatizovanih sistema za radiološku dijagnostifikaciju abnormalnosti tokom 2017. i 2018, radovi su se u većini slučajeva oslanjali na konvolucione neuronske mreže. One su obučavane u *multi-label* režimu koristeći binarnu unakrsnu entropiju kao funkciju gubitka. Rad (Wang et al., 2017) uveo je *ChestX-ray8* skup podataka, koji je kasnije proširen na 14 patoloških oznaka. Korišćenjem *DenseNet* i *ResNet* arhitektura *AUC* vrednosti su se po klasi kretale između 70 % i 84 %, u odnosu na patologiju. *CheXNet* (Rajpurkar et al., 2017) koristio je *DenseNet-121* i ostvario *AUC* od 76 % za klasu pneumonije na ovom skupu, tvrdeći da su te performanse slične onima koje bi radiolog imao. Na čelu sa istim istraživačem kao prethodno, (Rajpurkar et al., 2018) istraživao je polu-nadgledani pristup, kombinujući označene i neoznačene slike. Cilj je bio poboljšanje generalizacije kroz deljene reprezentacije (engl. *shared representation*) između klasa. Svim ovim radovima zajedničko je korišćenje jednog *backbone*-a za sve klase, i nezavisnim *sigmoid* izlazima po klasi, bez eksplicitnog modelovanja odnosa među abnormalnostima.

U periodu između 2019. i 2020. radi se na modeliranju nezvesnosti i korišćenju korelacija između klasa. Studija (Irvin et al., 2019) uvodi skup podataka *CheXpert* koji je odgovoran za eksplicitnu vrednost koja označava da radiolog nije siguran. Rad takođe ispituje različite strategije rukovanja labelama (ignorisanje, različite metode mapiranja itd.). *DenseNet* i *ResNet* bac-

kbone su ponovo prisutni, postizujući makro *AUC* vrednosti od 89 % na validacionom skupu. Prvobitno objavljen 2020, (Pham et al., 2021) pokušava da kroz hijerarhijsko modelovanje labela i *label smoothing* poboljša performanse na nivou *multi-label* klasifikacije. Koristeći *CheXpert* skup podataka, sa preko 20000 snimaka, zajedno sa ansambliranjem modela postiže *AUC* od 94 % na 5 izabranih patologija.

Od 2021. pa nadalje, interesovanje za polu-nadgledanim (engl. *semi-supervised*) i samo-nadgledanim (engl. *self-supervised*) paradigmatima jača. Interesovanje većinski je izazvano shvatanjem o neizvesnosti labela u javno dostupnim skupovima podataka. Iako (Rajpurkar et al., 2018) još te godine uvodi elemente polu-nadgledanog učenja, kasniji pristupi se oslanjaju na pretreniranje nad velikim količinama neoznačenih slika i transfer učenja ka specifičnim zadacima dijagnoze. (Tiu et al., 2022) je značajan pomak jer koristi *vision-language* pretreniranje, sa *CLIP-like* okvirom. Ovim omogućava *zero-shot* predikcije bez finog podešavanja na ciljnom skupu. Autori prijavljuju *AUC* vrednosti koje su konkurentne sa standardnim potpuno nadgledanim pristupima na *CheXpert* i *MIMIC-CXR* skupovima. Time je pokazan potencijal za izgradnju *foundation* modela u medicinskoj dijagnostici.

2022. godine događa se da transformeri postaju sve zastupljeniji, pa čak i *vision-language* foundation modeli. (Taslimi et al., 2022) koristi *Swin Transformer* arhitekturu kao glavni *backbone* na *NIH ChestX-ray14* skupu podataka, deli sliku na parčiće (engl. *patches*), i oni se obrađuju kroz hijerarhijske blokove *Swin Transformer-a*. Lokalni mehanizmi pažnje kroz pomeranje prozora (engl. *window shifting*) i *Multi-Layer Perceptron* na izlaznom sloju postižu prosečan *AUC* od 81 % što je konkurentno, i nešto bolje od *CNN* modela na istom skupu. Rad prikazuje da arhitekture koje se baziraju isključivo na transformer mrežama mogu biti konkurentne u odnosu na *CNN* modele.

U Tabeli 2. prikazan je skraćeni pregled prethodno opisanih radova.

Tabela 2. Pregled radova, pristupa i performansi

Rad	Pristup	Skup podataka	ROC AUC (%)
Wang et al., 2017	<i>Multi-label CNN (DenseNet / ResNet), BCE loss</i>	<i>NIH ChestX-ray14</i>	70-84 (po klasama)
Rajpurkar et al., 2017	<i>DenseNet-121, multi-label, fokus pneumonija</i>	<i>NIH ChestX-ray14</i>	76 (pneumonija)
Rajpurkar et al., 2018	Polu-nadgledani <i>CNN</i> , deljene reprezentacije	<i>NIH ChestX-ray14</i>	76-79
Irvin et al., 2019	<i>DenseNet / ResNet, multi-label, upravljanje „uncertain“ labelama</i>	<i>CheXpert</i>	89 (makro <i>AUC</i>)
Pham et al., 2021	Hijerarhijski <i>CNN</i> , <i>label smoothing</i> , ansambliranje modela	<i>CheXpert</i>	94 (5 selektovanih patologija)
Tiu et al., 2022	<i>Vision-language pre-training (CLIP-like), zero-shot</i>	<i>CheXpert, MIMIC-CXR</i>	~88-90 (po klasama)
Taslimi et al., 2022	<i>Swin Transformer, patch-based multi-label</i>	<i>NIH ChestX-ray14</i>	81 (prosek <i>AUC</i>)

4. POSTAVKA PROBLEMA I SKUP PODATAKA

U ovom poglavlju biće reči o problemima unutar analize rendgenskih snimaka, kao i problema modernih, javno dostupnih skupova podataka. Takođe će se predstaviti analiza skupa podataka *NIH ChestX-ray14*, kako bismo mogli znati koje korake treba preduzeti tokom obučavanja modela.

4.1. Postavka problema

U kliničkoj praksi radiografija, a posebno radiografija snimaka grudnog koša predstavlja jednu od najčešće korišćenih dijagnostičkih metoda. Za razliku od drugih slika koje se koriste u medicini, kao što su dermoskopske slike koje su visoko kontrolisane i fokusirane na regiju interesa, snimci grudnog koša predstavljaju projekcije trodimenzionalne anatomske strukture na dvodimenzionalnu ravan. Usled te projekcije, dolazimo do preklapanja različitih anatomske struktura (pluća, dijafragma, srce), što može otežati klasifikaciju abnormalnih obrazaca (Wang et al., 2017).

Priroda radiografskih snimaka takođe se vezuje za manjak kontrasta, kao i visoki dinamički opseg. Abnormalnosti kao što su infiltracija ili čvorovi, često se mogu manifestovati kao suptilne promene u gustini ili teksturi tkiva, predstavljajući problem u identifikaciji čak i iskusnijim radiolozima (Ropp et al., 2015). Kvalitet snimaka takođe ume varirati, uglavnom zbog različitosti u aparatima, ekspoziciji ili položaju pacijenta (Dasegowda et al., 2023).

Formirani su veliki skupovi podataka kao što su *NIH ChestX-ray14* i *Stanford University CheXpert*, omogućivši razvoj modela baziranim na različitim arhitekturama dubokih neuronskih mreža, u svrhu automatske analize radiografskih slika (Wang et al., 2017; Irvin et al., 2019). Nasuprot svojoj pretpostavljenoj raznovrsnosti, kao i veličini, ovi skupovi podataka ipak poseduju određena ograničenja. Naime, labela nisu anotirane ručno u ogromnom broju slučajeva, već su za njihovu ekstrakciju korišćene metode obrade prirodnog jezika (engl. *Natural Language Processing, NLP*) koje su automatski na osnovu izveštaja izvlačile prisutne klase. Ovo unosi potencijalni problem nesigurnosti labela, odnosno mogućnosti da je labela implicitno pomenuta, negirana ili ipak izražena sa dozom nesigurnosti.

U *CheXpert* skupu, kako bi se ta mogućnost greške ublažila, uveden je koncept „nesigurnih“ labela. Ovo znači da u specifičnim instancama radiografskih snimaka neke od labela mogu imati vrednost koja indikuje nesigurnost, dajući još jedan aspekt mogućnosti da se ovakve instance drugačije tretiraju, dajući dodatnu kompleksnost formulaciji problema.

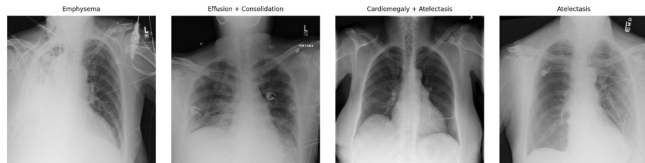
Medicinski problemi su karakteristični po neuravnoteženosti klasa (Mosquera et al., 2024; Roy et al., 2024). Neke klase su u kliničkim uslovima češće zastupljene od drugih. Takođe, postoji značajan broj slika koji u sebi sadrži istovremeno više prisutnih abnormalnosti. Tokom obučavanja modela, ova činjenica se može preneti na tendenciju da mreža implicitno uči korelacione obrasce, umesto učenja vizuelno distinktivnih karakteristika svake od patologija. Drugim rečima, modeli koji

su trenirani u klasičnoj *multi-label* postavci mogu naučiti da prisustvo jedne abnormalnosti implicira drugu, nasuprot činjenici da vizuelni aspekt to ne potvrđuje. Kod retkih klasa ovo postaje posebno problematično, gde se model može potpuno oslanjati na zavisnost od češćih, visoko korelisanih patologija.

4.2. Skupovi podataka

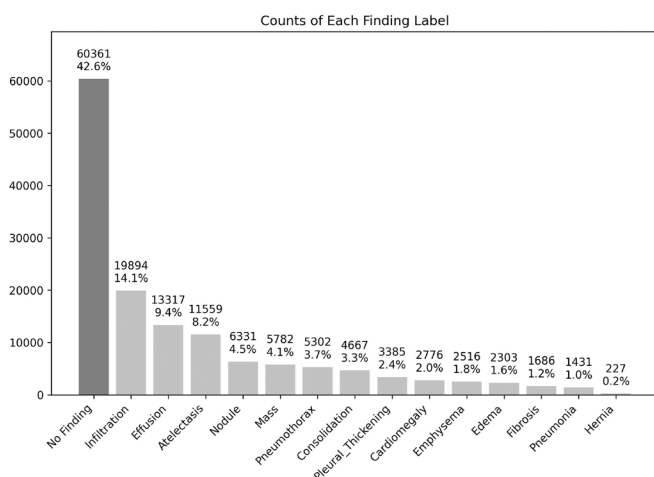
Prvi skup podataka, nad kojim će se izvršiti obučavanje, validacija i testiranje, je *NIH ChestX-ray14*, predstavljen 2017. godine. U poslednjih nekoliko godina dorađen je i verzija sa kojom ćemo raditi se bazira na skupu podataka sa *Kaggle-a*¹. Skup podataka sadrži 112.120 rendgenskih snimaka prikupljenih od 3805 pacijenata, čineći bazu poprilično velikom i raznovrsnom. Dimenzije imaju tendenciju da se razlikuju, širina se kreće od 1.143 *px* do 3.827 *px*, gde je prosek oko 2.646 *px*, a visina varira između 966 *px* i 4.715 *px*, sa prosekom na 2.486 *px*. Polovina slika ima širinu veću od 2.518 *px* i visinu veću od 2.544 *px*. Prisustvo velikog broja slika visoke rezolucije može doprineti boljoj reprezentaciji abnormalnosti. Skup sadrži veliku količinu slika, kombinuje rezolucije varirajućeg kvaliteta, što sve igra važnu ulogu u izgradnji generalizacije modela pri različitim uslovima.

Na Slici 1. prikazane su odabrane slike iz *NIH ChestX-ray14* skupa podataka.



Slika 1. Primeri rendgenskih snimaka iz *NIH ChestX-ray14* skupa podataka

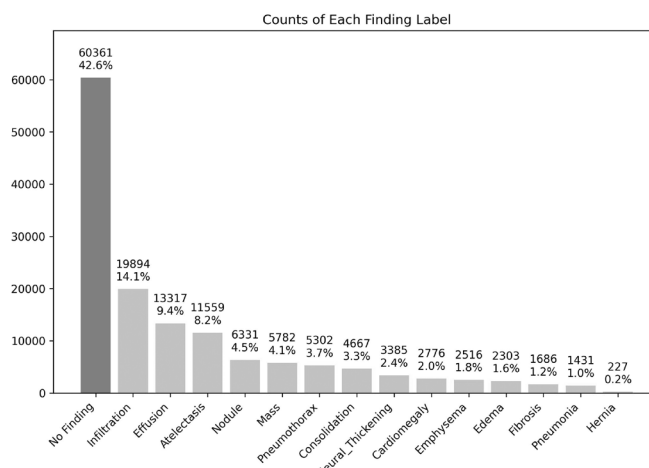
Raspodela klasa u skupu podataka prikazana je na Slici 2.



Slika 2. Balans klasa u *NIH ChestX-ray14* skupu podataka

Vidimo da 42,6 % skupa podataka predstavljaju slike bez ikakvih abnormalnosti, a takođe da su neke klase (*Fibrosis*, *Pneumonia* i posebno *Hernia*) mnogo slabije zastupljene od

drugih. Ova neuravnoteženost će uticati na pristup prilikom obučavanja modela, i takođe može ograničiti mogućnosti modela da dobro generalizuje. Uopšteni pristup, koji je primenjen u relevantnoj literaturi je baziran na tzv. *multi-label* principu, koristeći se jednom osnovom (engl. *backbone*) za obučavanje i učenje obrazaca za sve klase. Ovo je opravdano pretpostavkom o zajedničkom pojavljivanju klasa, kao i njihovom međusobnom povezanošću i zavisnošću. Na Slici 3. primetiće se međusobna povezanost različitih abnormalnosti u skupu podataka.



Slika 3. Povezanost klasa u skupu podataka

Primećujemo da *Atelectasis* ima veliku povezanost sa *Effusion*, *Infiltration* i *Consolidation* klasama, dok se klasa *Infiltration* često pojavljuje sa *Effusion*, *Nodule*, *Mass* i *Pneumothorax* klasama. Snažno se povezuju još *Consolidation* i *Effusion*, *Cardiomegaly* sa *Effusion* klasom, kao i *Emphysema* i *Pneumothorax*. Grafik ilustruje veliku međuzavisnost klasa.

Na Tabeli 3. prikazane su tačne korelacije između klasa. Za svaku od klasa prikazana je najzastupljenija koja se sa njom povezuje.

Tabela 3. Klase koje se najčešće pojavljuju zajedno

Abnormalnost 1	Abnormalnost 2	Broj zajedničkog pojavljivanja	Procenat zajedničkog pojavljivanja (%)
<i>Infiltration</i>	<i>Effusion</i>	4.000	20,1
<i>Edema</i>	<i>Infiltration</i>	981	42,6
<i>Atelectasis</i>	<i>Effusion</i>	3.275	28,3
<i>Pneumothorax</i>	<i>Effusion</i>	996	18,8
<i>Fibrosis</i>	<i>Infiltration</i>	345	20,5
<i>Mass</i>	<i>Effusion</i>	1.254	21,7
<i>Hernia</i>	<i>Atelectasis</i>	40	17,6
<i>Pneumonia</i>	<i>Infiltration</i>	605	42,3
<i>Consolidation</i>	<i>Effusion</i>	1.287	27,6
<i>Emphysema</i>	<i>Pneumothorax</i>	747	29,7
<i>Effusion</i>	<i>Infiltration</i>	4.000	30,0
<i>Pleural Thickening</i>	<i>Effusion</i>	849	25,1
<i>Nodule</i>	<i>Infiltration</i>	1.546	24,4

¹ Dostupno na: <https://www.kaggle.com/datasets/shivapan/nih-x-ray-dataset>

Vidimo da postoji veliki broj abnormalnosti koje se pojavljuju zajedno sa *Effusion* labelom. S druge strane, to je potpuno očekivano jer je *Effusion* često združeni simptom ili je izazvan istim primarnim uzorkom kao i niz drugih abnormalnosti. Primećujemo takođe, da je relativno veliki procenat zajedničkog pojavljivanja. Istražujući ovu činjenicu, nalazimo da u skupu podataka postoji 3.963 slike (27,6 %) koje imaju tačno jednu dijagnozu, dok više od jednog nalaza sadrži 2.796 (18,55 %) slika. Ukupno posmatrano, 51.759 slika sadrži barem jednu abnormalnost. Prosečan broj nalaza po slici (isključujući klasu *No Finding*) je 1.57, dok je maksimalan broj 9. Ovo nam može pomoći da razumemo manjak izolovanosti abnormalnosti unutar grudnog koša.

Skup podataka koji će se koristiti za unakrsno testiranje (engl. *cross-testing*) je *CheXpert*. Za ovaj skup je bitno ukazati na činjenicu da su labele manuelno ekstrahovane od strane eksperata, u odnosu na *NIH* skup koji se oslanja na automatsku ekstrakciju labela putem *NLP* metoda. Takođe je bitno pomenuti da su instance „nesigurnosti“ labelara konvertovane u klasu „neprisutan“ radi jasnoće. Unakrsno testiranje će biti izvedeno nad presekom labela *CheXpert* i skupa podataka za obučavanje, rezultujući u smanjenom broju labela. Manjak analize nad ovim skupom opravdan je idejom da on treba ostati istinski nezavisan, a da se analizom uvodi potencijalna pristrasnost.

5. EKSPERIMENTI I REZULTATI

Ovo poglavlje tiče se objašnjenja načina obučavanja modela na prethodno objašnjenom skupu podataka. Pojasniće se načini na koji je funkcija gubitka adaptirana, proces finog podešavanja, izabrane metrike i slično. Takođe će biti ilustrovani rezultati tokom i nakon obučavanja koji će se odnositi na odvojen skup podataka za testiranje.

5.1. Postavka eksperimenata

Kako bismo prevazišli probleme *multi-label* klasifikacije, primenjujemo binarni pristup posebno svakoj klasi. Binarni pristup otvara mogućnost boljoj adaptaciji praga odlučivanja, kao i olakšanu analizu rezultata. Rizici binarnog pristupa prvenstveno su vezani za duže vreme obučavanja, gde umesto obučavanja jedne osnove obučavamo model za svaku klasu. Pored toga, pošto obučavamo više osnova, otklanjamo mogućnost za prirodnom regularizacijom kroz učenje deljenih reprezentacija za sve. Efekat ove stavke ublažavamo korišćenjem *L2* regularizacije.

Binarni pristup nije karakterističan pristup problemu u literaturi, ali se koristi kako bismo mogli da svakoj abnormalnosti adaptiramo funkciju gubitka. Radi ublaženja neuravnoteženosti klasa koristi se *focal* gubitak (engl. *focal loss*), gde su koeficijenti prilagođavani u skladu sa zastupljenošću klasa. Ispod je prikazana formula.

$$\text{Focal Loss}(p_i) = -\alpha_i(1 - p_i)^{\gamma} \log(p_i)$$

gde je α faktor balansiranja klasa, a γ parametar fokusiranja. p predstavlja predviđenu verovatnoću ciljne klase. Faktor α je u većini slučajeva stavljen da bude 0,8 u svrhu stavljanja veće težine na manje zastupljene klase, ali je u slučaju klase *Pneumonia* stavljen na 0,9.

Optimalna kombinacija hiperparametara određena je mrežnom pretragom, uz trostruku stratifikovanu unakrsnu validaciju (Varoquaux, & Cheplygina, 2022). Konačan izbor je takođe baziran na prosečnim vrednostima kao i konvergenciji evaluacionih metrika kroz proces validacije.

Mreže su koristile pretrenirane težinske koeficijente sa *ImageNet* skupa, sa finim podešavanjem celokupne arhitekture. Primenjena je regularizacija (*L2* od $1e-4$), kako ne bi došlo do prenaučnosti, i u kombinaciji sa niskom stopom učenja $1e-5$ do $1e-4$ u zavisnosti od arhitekture i abnormalnosti (Goodfellow et al., 2016). Veličina serije je adaptirana u odnosu na memorijske mogućnosti sistema. *MobileNetV4* i *EdgeNeXt* se obučavaju duž 30 epoha, a *EfficientFormerV2* 20, sa mehanizmom ranog zaustavljanja nakon 10 epoha bez poboljšanja validacionog gubitka. Uzima se kontrolna tačka modela u odnosu na epohu kada je ostvaren najniži validacioni gubitak i taj se model koristi pri testiranju. Radi ostvarivanja bolje konvergencije modela, implementira se i smanjivanje stope učenja po platou validacionog gubitka, gde se na svakih 5 epoha ona smanjuje 10 puta.

Kako bi bila očuvana dijagnostička verodostojnost snimaka, augmentacije su ograničene na minimum. Korišćene su nasumična rotacija od $\pm 5^\circ$, u kombinaciji sa varijacijom oštine. One će pomoći u simuliranju manjih varijacija u položaju sonde, time obezbeđujući bolju generalizaciju.

Usled neuravnoteženosti skupa podataka, kao u literaturi, oslanjamo se na metrike nezavisne od praga odlučivanja, konkretno na površinu ispod *ROC* krive (engl. *Area Under the Receiver Operating Characteristic Curve*, *ROC AUC*). Ova metrika nam omogućava da dobijemo dobru ideju o diskriminaciji modela (Fawcett, 2006). Umesto binarne unakrsne entropije koja se koristi u većini radova, mi ćemo primeniti *focal* gubitak, i kroz modifikaciju težinskih koeficijenta na klasama (prisutan/neprisutan) asistiraćemo mreži da u boljem fokusiranju na primere koji su teži za klasifikaciju.

Obučavanje modela izvršeno je sa procesorom *AMD Ryzen 9 5950X*, grafičkom karticom *NVIDIA RTX 3090 (24 GB)*, 48 GB RAM-a (3200 MHz) i 1 TB M.2 NVMe ROM memorijom. Kod je dostupan na *GitHub*-u².

5.2. Rezultati

U Tabeli 4. prikazano je vreme obučavanja trostruke unakrsne validacije za svake od navedenih modela, maksimalna RAM zauzetost, kao i prosečno vreme zaključivanja po slici na CPU. Testiranje je izvršeno unutar *ONNX Runtime* sesija, kako bi se simuliralo izvršavanje u realnom okruženju.

² Dostupno na: <https://github.com/pipstur/cxr-report-generator>.

Tabela 4. Vreme obučavanja i prosečno vreme zaključivanja predstavljenih modela

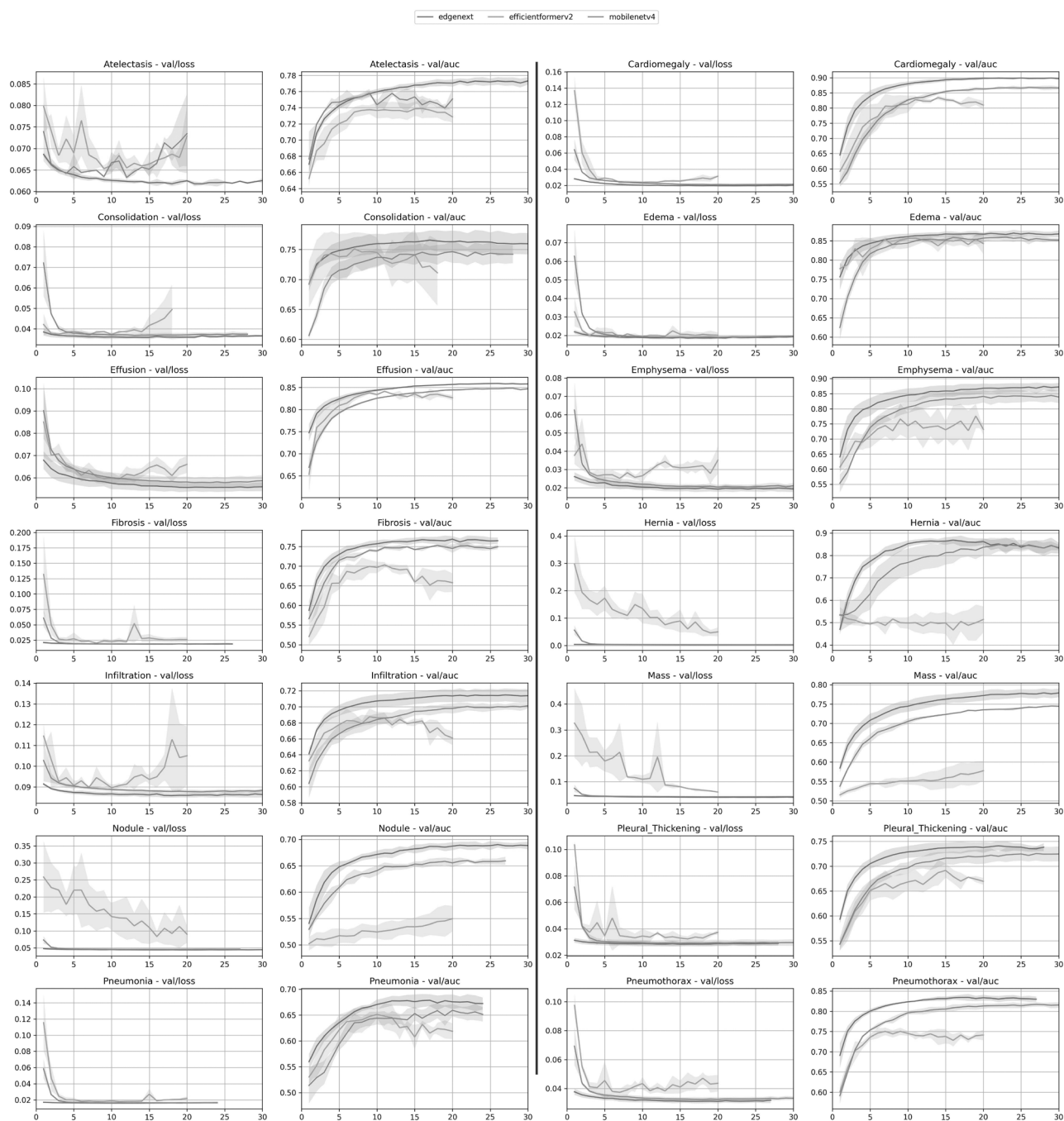
Model	Trajanje obučavanja (sati)	Maksimalna RAM zauzetost (MB)	Prosečno vreme zaključivanja na CPU (milisekunde po slici)
MobileNetV4	21	170	10.3
EfficientFormerV2	31,5	195	18,8
EdgeNeXt	42	154	22,5

Primećuje se da je *MobileNetV4* najbolji po pitanju vremena zaključivanja i obučavanja, ali da *EdgeNeXt* ipak zauzima manje memorijskog prostora. Bitno je napomenuti da je za obučavanje *EfficientFormerV2* modela potrebno 31,5 sati,

ali da je to većinski jer su modeli obučavani do 20 epoha, u odnosu na ostatak koji je obučavan 30. Ako bismo broj epoha izjednačili, vreme obučavanja *EfficientFormerV2* bi bilo slično kao od *EdgeNeXt* arhitekture.

Brzo zaključivanje i lagano okruženje za izvršavanje čini bilo koji od ovih modela pogodnim za izvršavanje na uređajima ograničenih računarskih mogućnosti. Vreme obučavanja nije dugo, te bi proces ponovnog obučavanja bio znatno lakši u odnosu na veće modele korišćene u literaturi.

Na Slici 4. prikazani su grafici koji predstavljaju konvergenciju prethodno predstavljenih modela, po pitanju validacionog gubitka i površine ispod ROC krive.



Slika 4. Konvergencija funkcije gubitka i ROC AUC prilikom validacije

Sa slike primećujemo da *MobileNetV4* i *EdgeNeXt* imaju dobru konvergenciju i po pitanju validacionog gubitka, kao i površine ispod ROC krive. *EfficientFormerV2*, sa druge strane, poseduje većinski nestabilniju i lošiju konvergenciju, pored lošije moći diskriminacije. Ove razlike posebno su očigledne ako posmatramo labele *Hernia*, *Nodule*, *Mass*, gde model duž celog procesa obučavanja ostaje loš po pitanju diskriminacione moći, indikovano po ROC AUC vrednostima koje su oko 50-55 %. *EdgeNeXt* postiže marginalno bolje rezultate od *MobileNetV4* arhitekture, po pitanju svake labele. Ovo može značiti da je ovaj model zbog suptilnog korišćenja mehanizma pažnje potencijalno pogodniji za rešavanje problema na rendgenskim slikama.

Lošiji rezultati ostvareni od strane *EfficientFormerV2* arhitekture su prvenstveno vezani za njenu veću osetljivost na izbor hiperparametara. Ovo bi se potencijalno moglo poboljšati detaljnijom optimizacijom. Loše performanse na klasama kao što su *Nodule* ili *Mass*, vezuju se za njegovo striktnije modeliranje globalnog konteksta. Ove klase više zavise od lokalnih struktura nego od globalnih paterna. *Hernia* ima veoma mali broj primera (0,2 % skupa podataka, 227 primera), je moguće da zbog toga model nije uspeo naučiti reprezentaciju. Takođe je moguće da bi ovoj arhitekturi više odgovarao *multi-label* pristup jer bi povećao količinu podataka i omogućio bolje hvatanje globalnog konteksta kroz bogatije deljenje reprezentacije klasa.

U literaturi, kao i u slučaju ovih modela, performanse su konzistentno najslabije po pitanju klase *Pneumonia* (Wang et al., 2017; Rajpurkar et al., 2017). Ova činjenica objašnjena je time što *Pneumonia* izgleda kao blago zamućenje koje, naročito kada je rezolucija smanjena, postaje teško za model da razlikuje od okolnog tkiva. Pored toga što je suptilna, u 42,3 % slučajeva se pojavljuje sa labelom *Infiltration*. Njena prisutnost je takođe među najnižima (1 % skupa podataka), tako da je potencijal za učenje dodatno smanjen.

Na Tabeli 5. prikazane su vrednosti površine ispod ROC krive, zajedno sa prilikom testiranja na odvojenom skupu podataka.

Tabela 5. ROC AUC vrednosti na NIH ChestX-ray14 skupu podataka, u procentima

Abnormalnost	<i>MobileNetV4</i>	<i>EfficientFormerV2</i>	<i>EdgeNeXt</i>
<i>Atelectasis</i>	76,63 ± 0,49	76,10 ± 0,56	77,72 ± 0,14
<i>Cardiomegaly</i>	88,21 ± 0,45	84,59 ± 1,38	90,02 ± 0,61
<i>Consolidation</i>	72,34 ± 0,34	73,01 ± 0,92	75,21 ± 0,07
<i>Edema</i>	86,41 ± 0,19	86,24 ± 0,51	87,39 ± 0,60
<i>Effusion</i>	84,95 ± 0,20	84,30 ± 0,32	85,92 ± 0,17
<i>Emphysema</i>	85,43 ± 0,46	77,83 ± 2,45	87,33 ± 0,29
<i>Fibrosis</i>	74,49 ± 0,80	69,83 ± 1,55	76,60 ± 0,79
<i>Hernia</i>	90,84 ± 0,43	52,67 ± 2,06	89,82 ± 0,87
<i>Infiltration</i>	69,60 ± 0,11	68,31 ± 0,08	70,65 ± 0,25
<i>Mass</i>	74,35 ± 1,45	58,34 ± 1,60	77,50 ± 0,31
<i>Nodule</i>	66,76 ± 0,58	54,71 ± 2,70	70,59 ± 0,67
<i>Pleural Thickening</i>	72,27 ± 0,20	68,70 ± 1,42	74,57 ± 0,58
<i>Pneumonia</i>	62,98 ± 1,37	62,27 ± 1,06	67,73 ± 1,25
<i>Pneumothorax</i>	82,90 ± 0,32	76,45 ± 0,99	83,20 ± 0,33
Prosek	77,73	70,95	79,59

Sa tabele primećuje se konzistentnost po boljim rezultatima ostvarenim od strane *EdgeNeXt* arhitekture, mada *MobileNetV4* ne zaostaje mnogo u većini slučajeva. *EfficientFormerV2*, doduše, jeste znatno lošiji od druga dva modela. Razlike su posebno izražene na *Nodule*, *Hernia* i *Mass* labelama. Ipak, *EfficientFormerV2* se pokazuje dobrim po pitanju diskriminacione sposobnosti kod *Cardiomegaly*, *Edema*, *Effusion*, *Emphysema* i *Pneumothorax*, što znači da ovaj model sadrži određenu vrednost. Moguće je da se pri drugačijem pristupanju procesu obučavanja rezultati poboljšaju i na drugim klasama. Svi modeli su znatno lošiji u detekciji *Pneumonia* labele, usled njene suptilnosti na rendgenskim snimcima.

Na Tabeli 6. nalaze se rezultati ostvareni prilikom unakrsnog testiranja na skupu podataka *CheXpert*. NIH prosek odnosi se na prosečnu diskriminaciju modela na podskupu klasa prilikom internog testiranja.

Tabela 6. ROC AUC vrednosti na CheXpert skupu podataka prilikom unakrsnog testiranja, u procentima

Abnormalnost	<i>MobileNetV4</i>	<i>EfficientFormerV2</i>	<i>EdgeNeXt</i>
<i>Atelectasis</i>	66,3	66,4	67,1
<i>Cardiomegaly</i>	75,1	71,5	78,2
<i>Consolidation</i>	65,1	66,6	66,3
<i>Edema</i>	64,2	74,9	74,2
<i>Effusion</i>	77,6	80,8	80,5
<i>Pneumonia</i>	58,5	53,3	56,5
<i>Pneumothorax</i>	67,6	60,8	69,5
Prosek	67,77	67,76	70,33
NIH prosek	79,20	77,57	81,03

Na osnovu prikazanih rezultata, uočava se pad performansi u odnosu na interni skup za testiranje, ali modeli prikazuju određenu dozu generalizacije. Izuzetak je klasa *Pneumonia*, gde se diskriminacija približava nasumičnim modelima. *EfficientFormerV2*, iako prethodno lošiji, prikazuje se kao dobar performer na ovom podskupu labele. Ipak, u proseku moći diskriminacije, i dalje *EdgeNeXt* ostaje najbolji. Prosečne vrednosti primetno padaju u slučaju svih arhitektura u odnosu na interno testiranje. Ovo je tipično u medicinskim zadacima gde razlika u opremi za snimanje, raspodelama klasa i procesu anotacija može napraviti drastičan pomeraj.

U kliničkim kontekstima najbolje bi bilo koristiti *EdgeNeXt* arhitekturu, usled superiornih rezultata po pitanju prosečnih vrednosti po abnormalnosti, ali i često manjoj devijaciji unutar unakrsne validacije. Ovaj model daje najbolje rezultate i na unakrsnom testiranju. Iako zahteva duže vreme za obučavanje, po pitanju brzine zaključivanja i memorijskog opterećenja sličan je ostalim modelima. Ovakav izbor modela obezbediće reprezentativne performanse u realnim uslovima.

6. ZAKLJUČAK

U radu je izložen pokušaj da se kompaktniji modeli koji koriste *CNN* i *ViT* paradigme obučavaju da detektuju abnormalnosti na rendgenskim snimcima grudnog koša. Mreže *MobileNetV4* i *EdgeNeXt* uspešno generalizuju prilikom obučavanja i testiranja, ostvarujući u proseku ROC AUC vrednosti od 77,73 % i 79,59 % na skupu podataka za testiranje, dok *EfficientFormerV2* ostvaruje 70,95 %. Ovo je u rang sa relevantnom lite-

raturu koja prikazuje 76-81 % prosečnih *ROC AUC* vrednosti, ali pri korišćenju znatno većih neuronskih mreža (*ResNet*, *DenseNet*). Unakrsnim testiranjem *AUC* vrednosti padaju za 9,81-11,43 % prikazujući pad u diskriminaciji, u očekivanim granicama. Rad ukazuje na moguće prednosti binarnog pristupa u kontekstu finog podešavanja pristupa svakoj od abnormalnosti, kao i potencijal za širenje i korišćenje mreža na uređajima sa ograničenim računarskim resursima.

Za budući rad predviđena je integracija komponente interpretabilnosti putem *Grad-CAM* slika, radi utvrđivanja tačnih regija slike na koje se modeli fokusiraju. Takođe će biti primenjene različite tehnike podešavanja granice odlučivanja, zajedno sa izgradnjom robusnijeg pretprocesiranja. Uz veći fokus na proces optimizacije, trud će biti uložen i u poboljšanje performansi modela na klasi *Pneumonia*. U daljoj budućnosti modeli se mogu integrisati sa velikim jezičkim modelima koji bi služili u automatskom generisanju izveštaja, dodatno umanjujući opterećenost radiologa ili doktora.

REFERENCE

- [1] Amin, S. U., & Hossain, M. S. (2020). Edge intelligence and Internet of Things in healthcare: A survey. *IEEE Access*, 9, 45-59. https://www.researchgate.net/publication/347641784_Edge_Intelligence_and_Internet_of_Things_in_Healthcare_A_Survey
- [2] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends® in Finance*, 2(1), 1-127. https://deeplearning.lipinyang.org/wp-content/uploads/2016/12/Learning-Deep-Architectures-for-AI_Yoshua-Bengio.pdf
- [3] Dasgowda, G., Kalra, M. K., Abi-Ghanem, A. S., Arru, C. D., Bernardo, M., Saba, L., ... & Dreyer, K. J. (2023). Suboptimal chest radiography and artificial intelligence: the problem and the solution. *Diagnostics*, 13(3), 412. <https://pubmed.ncbi.nlm.nih.gov/36766516/>
- [4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. <https://arxiv.org/pdf/2010.11929>
- [5] European Respiratory Society. (2023). The burden of lung disease. Retrieved from <https://www.ersnet.org/wp-content/uploads/2023/01/Overview.pdf>
- [6] Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874. <https://www.math.ucdavis.edu/~saito/data/roc/fawcett-roc.pdf>
- [7] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2, pp. 1-800). Cambridge: MIT press. <https://synapse.koreamed.org/pdf/10.4258/hir.2016.22.4.351>
- [8] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., ... Le, Q. (2019). Searching for MobileNetV3. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). http://openaccess.thecvf.com/content_ICCV_2019/papers/Howard_Searching_for_MobileNetV3_ICCV_2019_paper.pdf
- [9] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597). <https://aaai.org/ojs/index.php/AAAI/article/view/3834/3712>
- [10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://hal.science/hal-04206682/document>
- [11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2002). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://hal.science/hal-03926082/document>
- [12] Li, Y., Hu, J., Wen, Y., Evangelidis, G., Salahi, K., Wang, Y., ... & Ren, J. (2023). Rethinking vision transformers for mobilenet size and speed. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 16889-16900). https://openaccess.thecvf.com/content/ICCV2023/papers/Li_Rethinking_Vision_Transformers_for_MobileNet_Size_and_Speed_ICCV_2023_paper.pdf
- [13] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. ICCV. http://openaccess.thecvf.com/content/ICCV2021/papers/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.pdf
- [14] Li, Y., Yuan, G., Wen, Y., Hu, J., Evangelidis, G., Tulyakov, S., ... & Ren, J. (2022). Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems*, 35, 12934-12949. https://proceedings.neurips.cc/paper_files/paper/2022/file/5452ad8ee6ea6e7dc41db1c6d31ba0b8-Paper-Conference.pdf
- [15] Kasalak, Ö., Alnahwi, H., Toxopeus, R., Pennings, J. P., Yakar, D., & Kwee, T. C. (2023). Work overload and diagnostic errors in radiology. *European Journal of Radiology*, 167, 111032. <https://www.sciencedirect.com/science/article/pii/S0720048X23003467>
- [16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25. <https://proceedings.neurips.cc/paper/2012/file/c399862d-3b9d6b76c8436e924a68c45b-Paper.pdf>
- [17] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. <https://arxiv.org/pdf/1702.05747>
- [18] Maaz, M., Shaker, A., Cholakkal, H., Khan, S., Zamir, S. W., Anwer, R. M., & Shahbaz Khan, F. (2022, October). Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications. In European conference on computer vision (pp. 3-20). Cham: Springer Nature Switzerland. <https://arxiv.org/pdf/2206.10589>
- [19] Mitchell C. (2012). World radiography day: Two-thirds of the world's population has no access to diagnostic imaging. Pan American Health Organization.
- [20] Mosquera, C., Ferrer, L., Milone, D. H., Luna, D., & Ferrante, E. (2024). Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance. *European Radiology*, 34(12), 7895-7903. <https://drive.google.com/file/d/1P48bLV0m-2Z8IXPJxI5U-eHUuSI-Jyez/view>
- [21] Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T., & Nguyen, H. Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*, 437, 186-194. <https://arxiv.org/pdf/1911.06475>
- [22] Qin, D., Lechner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., ... & Howard, A. (2024, September). MobileNetV4: Universal models for the mobile ecosystem. In *European conference on computer vision* (pp. 78-96). Cham: Springer Nature Switzerland. <https://arxiv.org/pdf/2404.10518>
- [23] Roy, D., Roy, A., & Roy, U. (2024). Learning from imbalanced data in healthcare: State-of-the-art and research challenges. *Computational Intelligence in Healthcare Informatics*, 19-32. https://link.springer.com/chapter/10.1007/978-981-99-8853-2_2
- [24] Ropp, A., Waite, S., Reede, D., & Patel, J. (2015). Did I miss that: subtle and commonly missed findings on chest radiographs. *Current problems in diagnostic radiology*, 44(3), 277-289. <https://www.sciencedirect.com/science/article/abs/pii/S0363018814001157>
- [25] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content_cvpr_2018/papers/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.pdf
- [26] Taslimi, S., Taslimi, S., Fathi, N., Salehi, M., & Rohban, M. H. (2022). Swinhex: Multi-label classification on chest x-ray images with transformers. *arXiv preprint arXiv:2206.04246*. <https://arxiv.org/pdf/2206.04246>
- [27] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1), 44-56. https://shilonglenys.com/wp-content/uploads/2025/07/Eric-J-Topol-high-performance_medicine_the_convergence_of_human_and_artificial_intelligence_nature_medicine_volume_25_january_2019.pdf
- [28] Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digital Medicine*, 5(1), 48. <https://www.nature.com/articles/s41746-022-00592-y.pdf>
- [29] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX ray8: Hospital scale chest X ray database and benchmarks on weakly supervised classification and localization of common thorax diseases. *ArXiv Preprint*. http://openaccess.thecvf.com/content_cvpr_2017/papers/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.pdf



Vojislav Stevanović, Machine Learning Engineer u Darwin Digital d.o.o

Kontakt: vojislavstevanovic171@gmail.com

Oblasti interesovanja: kompjuterska vizija, neuronske mreže, procesiranja signala, nauka o podacima



Pavle Milošević, Vanredni profesor na Fakultetu organizacionih nauka u Beogradu

Kontakt: pavle.milosevic@fon.bg.ac.rs

Oblasti interesovanja: kompjuterska vizija, biometrija, fazi logika, teorija sistema