

Класификација медицинских слика плућа за рано откривање пнеумоније

Милош Чутурић - R2 13/2024

Лука Ђорђевић - R2 18/2024

Марко Јаношевић - R2 16/2024

Софтверско инжењерство и информационе технологије
Универзитет у Новом Саду, Факултет техничких наука

Нови Сад, Србија

I. Проблем и мотивација

Пнеумонија је један од најзначајнијих узрока смртности од инфективних болести у свету и водећи узрок смрти код деце млађе од пет година. Према подацима Светске здравствене организације и УНИЦЕФ-а, сваке године од пнеумоније умре више стотина хиљада деце, што представља око 14–15% свих смртних исхода у овој старосној групи. [1] Осим деце, посебно су угрожени старији пациенти и особе са хроничним оболељењима или ослабљеним имуним системом. Пнеумонија је, дакле, и даље глобални здравствени проблем упркос напретку у вакцинацији, антибиотској терапији и интензивној нези [2].

Са клиничке стране, благовремена дијагноза пнеумоније је кључна за исход лечења. Одлагање у постављању дијагнозе или погрешна интерпретација налаза доводе до кашњења у терапији, компликација и већег ризика смртног исхода. Стандардни дијагностички приступ заснива се на клиничкој слици (повишене температура, кашаљ, отежано дисање, бол у грудима) и дијагностичким методама као што су лабораторијске анализе и радиолошко снимање. Иако се компјутеризована томографија сматра „златним стандардом“ за овај проблем, у пракси се најчешће користе рендгенски снимци грудног коша због ниže цене, доступности и никег зрачења [3].

Међутим, интерпретација рендгенских снимака плућа захтева велико искуство радиолога. Карактеристичне промене код пнеумоније (инфилтрати, замућења) могу бити суптилне, преклапати се са другим патологијама или бити одсутне у раним фазама болести. Ово отвара простор за субјективност и људске грешке, посебно у условима велике оптерећености здравственог система, када један радиолог анализира велики број снимака у кратком времену. Поред тога, у многим сиромашнијим или руралним срединама уопште немаово обучених радиолога, па дијагноза често зависи од једног лекара опште праксе, што додатно повећава ризик од погрешне или закаснеле дијагностике [3].

Горенаведене чињенице доводе до закључка да су компјутерски асистирани дијагностички системи неопходни као допуна постојећој пракси.

II. Теоријске основе

A. Конволуционе неуронске мреже

Конволуционе неуронске мреже (CNN) представљају основни приступ у анализи медицинских снимака јер омогућавају ефикасно издавање просторних образца и текстуралних промена применом конволуционих филтера који имитирају начин на који визуелни кортекс детектује локалне структуре [4]. Управо зато су CNN мреже постале доминантне у анализи рендгенских снимака, укључујући детекцију пнеумоније, што је потврђено и у CheXNet раду [5].

Трансфер учење омогућава употребу великог знања „уграђеног“ у велике преттрениране моделе, што је посебно важно у медицини где су склопови података ограничени [6]. У овом раду коришћене су архитектуре трениране на ImageNet склопу података, DenseNet121, ResNet50 и EfficientNetB0 због своје доказане ефикасности у класификацији медицинских снимака:

- **DenseNet121** - користи густо повезане слојеве како би олакшао проток градијента и реупотребу карактеристика [7].
- **ResNet50** - уводи резидуалне скокове који стабилизују тренинг веома дубоких мрежа [8].
- **EfficientNetB0** - примењује compound scaling приступ који оптимално усклађује дубину, ширину и резолуцију модела [9].

Додатно је коришћен и **CheXNet** који представља варијанту DenseNet121 преттренирану на великом ChestX-ray14 склопу са више од 100.000 снимака грудног коша, специјално оптимизована за детекцију плућних оболења [5].

На све моделе додати су Dense слојеви са Batch-Norm и Dropout регуларизацијом како би се надоградиле високонивоовске карактеристике преттренираних мрежа и смањио ризик пренаглашених учења при бинарној класификацији [10].

B. Генеративне адверзеријалне мреже

Обзиром да је коришћени склоп података небалансиран (поглавље Скуп података IV), било је неопходно генерисати додатне податке помоћу генеративне адверзеријалне мреже (GAN). Ове мреже користе

интеракцију генератора и дискриминатора како би научиле дистрибуцију података и генерисале нове примере, што их чини погодним за аугментацију у случајевима неуравнотежених података [11]. Генератор има задатак да на основу случајног шума производи синтетичке слике које имитирају реалне рендгенске снимке, док дискриминатор процењује да ли је слика реална или генерисана и тиме пружа сигнал генератору како да побољша своје излазе. Овај конкурентни процес омогућава постепено приближавање генерисаних слика стварној дистрибуцији података. У овом раду примењује се WGAN (Wasserstein GAN), који уводи Wasserstein-1 растојање као метрику разлике између реалне и генерисане дистрибуције, што резултује стабилнијим тренингом и реалистичнијим сликама у односу на класични GAN [12]. У медицинским задацима, укључујући анализу рендгенских снимака, WGAN је показан као ефикасан алат за генерисање синтетичких примера малобројних класа, чиме се умањује проблем неуравнотежености и побољшава робусност модела [13].

C. Метрике

За евалуацију модела коришћене су метрике релевантне за медицинску дијагностику:

- **Precision и Recall** - мере тачност позитивних предвиђања и способност модела да открије све случајеве болести [14].
- **F1-score** - обезбеђује уравнотежену процену ове две метрике, што је важно у неуравнотеженим скуповима
- **AUC** - мери укупну дискриминативну способност модела независно од прага одлучивања, што је стандард у медицинској машинској дијагностици [15]

III. Сродни радови

A. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Један од најзначајнијих радова у области аутоматске детекције пнеумоније је CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning [5]. Аутори у овом раду представљају модел DenseNet121 трениран на великим ChestX-ray14 скупу који садржи преко 100.000 рендгенских снимака са 14 класа оболења. Модел је показао перформансе упоредиве са радиолозима у детекцији пнеумоније, што је први пут демонстрирало да дубоко учење може достићи клинички ниво у анализи рендгенских снимака. У овом раду овај модел представља полазну основу. Користи се иста архитектура (DenseNet121), али је додатно прилагођена бинарном проблему (NORMAL/PNEUMONIA) кроз додавање Dense слојева и fine-tuning само појединих дубљих слојева. Главна разлика је у величини скупа података. Одобрани скуп података је знатно мањи, због чега не тренирамо модел од нуле већ примењујемо трансфер учење и агресивније технике регуларизације.

Такође, за разлику од CheXNet-а, уведена је GAN-базирана аугментација како би се елиминисао проблем неуравнотежених класа.

B. Pneumonia detection using transfer learning

Други релевантан рад је Pneumonia Detection Using Transfer Learning [16]. У наведеном раду истражује се примена трансфер учења на истом Kaggle скупу података који се користи и у овом раду, а тестирано је више преттренираних модела (VGG16, VGG19, ResNet50). Аутори показују да трансфер учење значајно побољшава тачност у односу на тренинг модела од нуле, посебно због ограничених величине и неуравнотежености скупа. Овај рад надограђује наведени приступ увођењем напреднијих архитектура (DenseNet121, EfficientNetB0), као и CheXNet-а који је специјално трениран на медицинским снимцима, што рад из спецификације није обухватио.

C. A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks

Трећи рад, A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks [3], представља систематски преглед метода коришћених за детекцију пнеумоније, укључујући CNN моделе, трансфер учење и различите технике аугментације. Аутори закључују да је трансфер учење тренутно најпоузданiji приступ због недовољне количине медицинских података, али истичу и кључне изазове, пре свега неуравнотежене класе, ризик од пренаглашеног учења и велики варијабилитет квалитета снимака. Наведени рад доприноси овом пројекту тако што идентификује ове проблеме и теоријски оправдава примену напредних техника као што су GAN-аугментација и fine-tuning дубоких модела. Овај рад директно одговара на проблеме описане у прегледу:

- трансфер учење је примарни приступ
- примењују се регуларизационе технике и специфични слојеви како би се избегло пренаглашено учење на малом скупу података

IV. Скуп података

У овом раду коришћен је јавно доступан скуп података Chest X-Ray Images (Pneumonia) који је објавио Mooney на Kaggle платформи [17]. Скуп садржи укупно 5.863 рендгенских снимака грудног коша у JPEG формату, подељених у две класе: NORMAL и PNEUMONIA. Снимци потичу из педијатријске клинике (Guangzhou Women and Children's Medical Center) и обухватају узраст 1–5 година.

Скуп података је организован у три засебне целине: train, test и validation, при чему је расподела по класама значајно неуравнотежена, што представља један од главних изазова. Табела I представља детаљну структуру коришћеног скупа података.

Главни изазов овог скупа података јесте значајна неуравнотеженост класа, јер је број снимака пнеумоније

Табела I
Структура Chest X-Ray датасета

Подскуп	NORMAL	PNEUMONIA	Укупно
Train	1341	3875	5216
Test	234	390	624
Validation	8	8	16
Укупно	1583	4273	5856

приближно три пута већи од броја нормалних снимака. Оваква дистрибуција може довести до тога да модел фаворизује већинску класу, што у медицинским задацима резултује опасним смањењем осетљивости за нормалне случајеве или, обрнуто, повећањем лажно позитивних детекција. Поред тога, скуп података обухвата слике различитих резолуција, осветљења и степена контраста, што може негативно утицати на стабилност тренинга и способност модела да научи униформне карактеристике.

Имајући у виду наведене варијабилности података, као и потребу за униформним улазом у дубоке моделе, над скупом података је спроведен низ корака претпроцесирања. Ови кораци омогућавају да се слике доведу у доследан формат погодан за тренинг CNN и трансфер learning модела, као и да се на одговарајући начин припреми балансирање између класа. Детаљан опис примењених техника, укључујући скалирање, нормализацију, аугментацију и поступке синтетичког проширења података, биће изложен у наредном поглављу.

V. Методологија

A. Претпроцесирање

Све слике се приликом учитавања доводе на фиксну величину 224×224 пиксела. Ово је потребно како би модели DenseNet121, ResNet50 и EfficientNetB0 добили улаз који је компатибилан са њиховом оригиналном ImageNet конфигурацијом. Слике се учитавају у RGB формату, иако су рендгенски снимци по својој природи сиви, јер преттренирани модели очекују три канала. У пракси се исти сиви интензитет копира у сва три канала.

Нормализација интензитета се врши тако што се све вредности пиксела скалирају у опсег $[0,1]$. На тај начин се стабилизује тренинг, избегавајујући сувише велике вредности активација у дубоким мрежама и убрзавајући конвергенцију.

Главни проблем Chest X-Ray скупа података јесте изразита неуравнотеженост класа. Број снимака класе PNEUMONIA је приближно три пута већи од броја NORMAL снимака. Оваква диспропорција доводи до тога да модел током тренинга развија пристрасност ка већинској класи, што резултује лошим перформансама на мањинској класи и значајно нарушава клиничку употребљивост модела.

Да би се овај проблем ублажио, модул спроводи класичан oversampling мањинске класе. Циљ је да се број NORMAL примера повећа све док не достигне број PNEUMONIA примера.

Повећање броја снимака реализује се генерирањем нових NORMAL примера применом аугментације, при чему се постојећи NORMAL снимци насумично бирају и над њима се примењују трансформације:

- **Ротирање** - слика се ротира за мали угао, чиме се симулирају варијације у позиционирању пацијента и рендген апарате
- **Трансляције по ширини и висини** - модел чине мање осетљивим на то да ли је грудни кош тачно центриран на снимку
- **Зумирање** - модел учи да препозна исте обрасце и када су структуре ближе или даље
- **Смицање** - како би се симулирале типичне неправилности које се јављају при снимању
- **Хоризонтално флиповање** - окретање слике по хоризонтали, што је прихватљиво за рендген плоча јер не наруша дијагностички значај
- **Попуњавање празних пиксела** - врши се тако да се новоформиране области попуњавају најближим постојећим вредностима, без увођења вештачких структура.

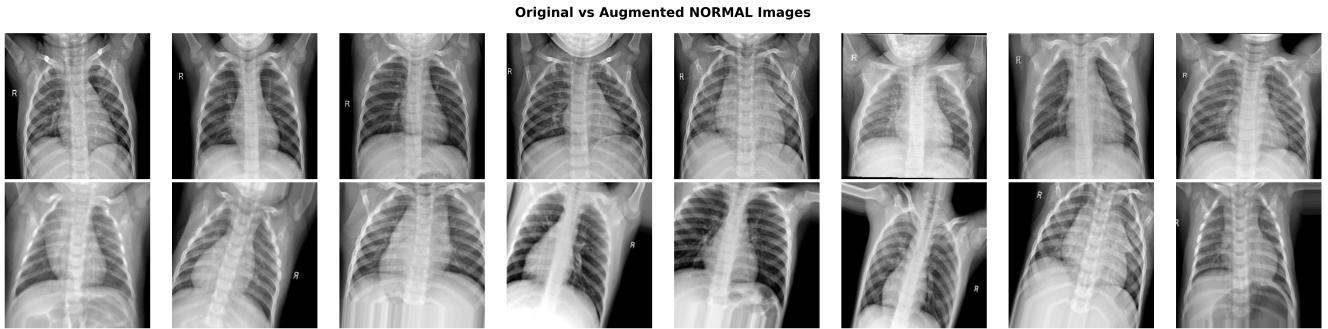
Слика 1 приказује пример снимака пре и после аугментације.

B. Допуна валидационог скупа - GAN

Оригинални валидациони скуп у Chest X-Ray скупу података садржи свега 16 слика (8 NORMAL и 8 PNEUMONIA), што је недовољно за стабилну процену перформанси модела и поуздано праћење напретка током тренинга. Већ једна погрешна класификација значајно мења вредности метрика, па је тешко проценити да ли модел заиста напредује или се пренаглашава на мали број примера. Због тога је уведен GAN модел, чији је циљ да генерише додатне синтетичке слике и тиме прошири валидациони скуп, уз задржавање исте класне структуре.

У овом пројекту је одабран WGAN (Wasserstein GAN), који се састоји од два дела: генератора и дискриминатора. Генератор прима случајни вектор и постепеним увећањем просторне резолуције кроз низ слојева гради слику рендгена, почевши од мале „латентне“ представе па до финалне слике одговарајуће величине. Дискриминатор, са друге стране, прима слику и учи да јој додели што реалистичнију оцену реалности. У WGAN формулацији дискриминатор не даје вероватноћу класе већ скаларну вредност, а разлика између просечних оцена за реалне и генерисане слике приближава Wasserstein-1 растојање између дистрибуција, што омогућава стабилнији тренинг у односу на класични GAN. Опробан је и класичан GAN модел, али није давао адекватне резултате.

Тренинг модела организован је тако да се у сваком кораку прво више пута ажурира дискриминатор, како би



Слика 1. Снимци пре и после аугментације

добро разликовао реалне и генерисане слике, а затим се ажурира генератор са циљем да превари дискриминатор и произведе слике које изгледају што реалније. Слике из тренинг скупа се претходно скалирају на низу резолуцију погодну за GAN, нормализују у опсег који одговара Tanh активацији генератора, а затим се користе као реални примери током учења. Након конвергенције, обучени генератор може да ствара нове синтетичке слике који визуелно подсећају на полазне рендгенске слике.

Након тренинга, модел се користи за генерисање додатних валидационих слика за обе класе. Генерисане слике се по потреби рескалирају на 224×224 пиксела и форматирају тако да буду компатибилне са истим претпроцесирањем као и остале слике. У табели II може се видети коначан скуп података који ће се користити за тренирање модела.

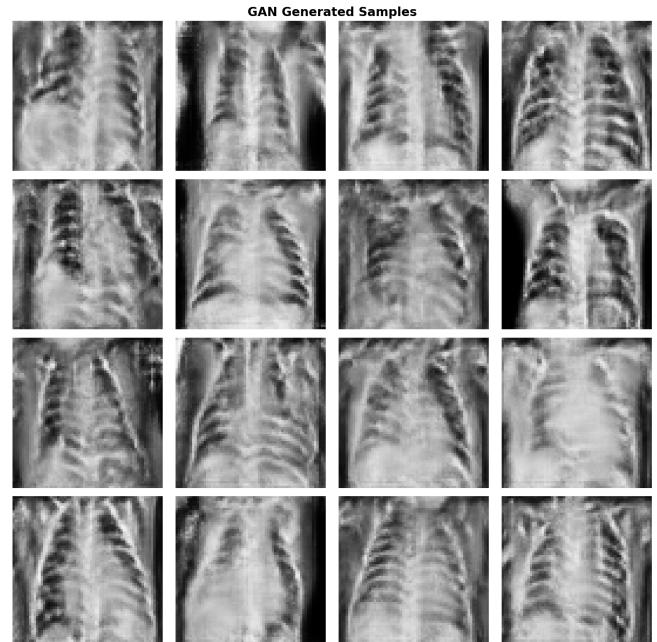
Табела II
Финални скуп података

Подскуп	NORMAL	PNEUMONIA	Укупно
Train	3875	3875	7750
Test	234	390	624
Validation	400	400	800
Укупно	4509	4665	9174

C. Конфигурација и тренинг модела

Базне архитектуре које се користе су DenseNet121, ResNet50 и EfficientNetB0, као и варијанта DenseNet121 специфично преттренирана за медицинске рендгенске снимке (CheXNet), као што је описано у поглављу Теоријске основе II. Све ове архитектуре учитане су без завршног класификатора и са улазном димензијом $224 \times 224 \times 3$, што је стандардни формат за ImageNet и уједно формат на који су претпроцесиране све слике у датом систему.

На излаз базног модела додаје се Global Average Pooling слој. Уместо да се feature map flatten-ује у велики вектор, глобално просечно пуловање израчунава по једну вредност по каналу (просек активација), чиме се драстично смањује број параметара и ризик од пренаглашеног учења.



Слика 2. Синтетички примери генерисани уз помоћ GAN-а

Након Global Average Pooling-а, дефинисана је следећа структура класификационог head-а:

- **Batch Normalization** - нормализација излаза из pooling слоја стабилизује тренинг и омогућава употребу релативно већих стопа учења [18]
- **Dense слој са 512 јединица и ReLU активацијом** - овај слој служи да из карактеристика извучених из базног модела формира богатију представу специфичну за задатак бинарне класификације NORMAL/PNEUMONIA. Вредност 512 представља компромис између изражажности модела и ризика од overfitting-а на релативно малом скупу података [19]
- **Batch Normalization и Dropout (стопа око 0.5)** - BatchNorm даље стабилизује дистрибуцију активација, а Dropout са релативно високом стопом служи као снажан регуларизатор, што је посебно важно код модела са преттренираним језгром и малим бројем узорака [10]

- **Dense слој са 256 јединица и ReLU активацијом** - овај слој додатно усавршава представу и омогућава моделирање сложенијих нелинеарних односа између extracted feature-а и коначне класе. Смањење броја неурона са 512 на 256 представља постепену редукцију димензијалности и додатно делује регуларизационо [20]
- **Batch Normalization и Dropout (стопа око 0.3)** - пошто је број параметара мањи него у претходном Dense слоју, dropout стопа се благо смањује, задржавајући регуларизациони ефекат без претераног гашења неурона [10]
- **Dense слој са 2 јединице и Softmax активацијом** - овај слој даје вероватноће за две класе. Softmax са две класе омогућава симетричну интерпретацију у оквиру категоријалног cross-entropy loss-а [20]

Оваква архитектура head-а директно се надовезује на препоруке из прегледног рада Alapat et al. да се CNN/transfer learning модели за пнеумонију надограде додатним густим слојевима и снажном регуларизацијом, како би се прилагодили специфичностима конкретног скупа података [3].

Као оптимизатор примењује се Adam са малом почетном стопом учења реда 10^{-4} током иницијалног тренинга класификацијоног дела модела и још низом вредношћу, приближно 10^{-5} , у фази fine-tuning-а дубљих слојева преттренираних мрежа. Adam је изабран јер адаптивно прилагођава стопу учења по параметру, што га чини изузетно стабилним и погодним за ситуације са шумовитим градијентима и малим скуповима података [21]. Као функција губитка користи се категоријални cross-entropy, што је стандардни приступ када се користи Softmax на излазу [5]. За евалуацију модела прате се метрике Precision, Recall, F1-score и AUC, јер прегледи литературе показују да у медицинским задацима класична тачност није довољна метрика. Неопходно је мерити и осетљивост, тј. способност отварања болести и дискриминативну моћ независну од прага класификације [3].

У почетној фази тренинга, базни модели се користе као екстрактори карактеристика. Њихови конволуциони слојеви држе се замрзнутим, тј. неажурирају се тежине, а тренира се само новододати класификациони део, head. Ово је стандардна стратегија у transfer learning приступима за медицинске слике, јер спречава драстичне промене већ добро научених карактеристика на малом доменском скупу [3].

Након иницијалног тренинга head-а, у другој фази се одмрзва мањи број последњих слојева базног модела и врши се fine-tuning са низом стопом учења. Fine-tuning последњих блокова значајно поправља перформансе на доменски специфичним подацима, као што су рендгенски снимци плућа [22].

D. Основни CNN модел

Поред преттренираних архитектура, у раду је имплементиран и baseline CNN модел. Циљ овог

модела јесте да представља контролни експеримент, односно полазну тачку за поређење са transfer-learning приступима. Модел се тренира од нуле како би се показало колико предзнање преттренираних мрежа доприноси коначним перформансама на малим медицинским скуповима. На тај начин, овај модел служи као референтна тачка која омогућава објективно мерење добитака постигнутих употребом напреднијих архитектура као што су DenseNet121, ResNet50 или EfficientNetB0.

Архитектура основног модела заснива се на класичној секвенци конволуционих слојева, pooling операција и потпуно повезаних (Dense) слојева. Мрежа почиње са конволуционим слојевима који користе мањи број филтера, што је уобичајен приступ код модела који се тренирају од нуле на ограниченом броју медицинских снимака. Сваки конволуциони блок прати активација ReLU и MaxPooling, којом се постепено смањује просторна резолуција и извлаче све апстрактније карактеристике. Оваква комбинација слојева представља традиционални CNN дизајн који се редовно примењује у задацима класификације радиолошких снимака када није доступан велики тренинг скуп [3].

Након конволуционих блокова, модел користи Flatten слој како би се просторне карактеристике претвориле у линеарни вектор, а Dense слој са Relu активацијом који служи као класификацијони део мреже. Присутан је и Dropout, који има важну улогу у спречавању пренаглашеног учења, посебно када се модел тренира од нуле на релативно малом скупу података. Финални слој је Dense слој са две јединице и Softmax активацијом, што омогућава излаз у виду вероватноћа за класе NORMAL и PNEUMONIA. Овакав дизајн омогућава да се јасно уочи колико боље перформирају дубље архитектуре са преттренираним тежинама. Зато основни CNN у овом раду функционише као контролни модел који даје ниже границу перформанси и показује оправданост примене трансфер учења на рендгенским снимцима плућа.

VI. Резултати и дискусија

Перформансе модела оцењене су на независном тест скупу који садржи 624 рендгенска снимка, а коришћене метрике укључују Precision, Recall, F1-score и AUC. Ове метрике су нарочито релевантне за медицинску дијагностику, јер литература указује да се accuracy не сме посматрати изоловано, већ у контексту осетљивости и специфичности модела [3]. У табели III могу се видети резултати тренираних модела на тест скупу, а у табели IV параметри одабрани за сваки модел.

A. DenseNet121

DenseNet121 постиже уравнотежен однос осетљивости и прецизности, са укупном тачношћу од 80.6% и AUC вредношћу од 0.8996 на обе класе. PNEUMONIA класа показује високу осетљивост (Recall = 0.9436), што је критично за клиничке примене, док је Recall за NORMAL класу осредњи (0.5769), што указује да модел има

Табела III
Перформанс модела на тест скупу

Модел	Accuracy	Precision (N/P)	Recall (N/P)	F1 (N/P)	Support (N/P)	AUC
DenseNet121	0.8061	0.6943 / 0.8626	0.5769 / 0.9436	0.5949 / 0.8588	260 / 364	0.8996
ResNet50	0.7397	0.5218 / 0.9187	0.3120 / 0.9949	0.3381 / 0.9555	250 / 374	0.9294
EffNetB0	0.5962	0.3262 / 0.7287	0.3077 / 0.7371	0.2773 / 0.7318	260 / 364	0.6254
BaselineCNN	0.7452	0.5440 / 0.9083	0.3333 / 0.9923	0.4091 / 0.8296	255 / 369	0.9367
CheXNet	0.7644	0.5721 / 0.8896	0.4274 / 0.9667	0.4881 / 0.9260	248 / 376	0.8610

Табела IV
Параметри и конфигурација коришћених модела

Модел	Параметри и конфигурација
DenseNet121	<ul style="list-style-type: none"> Преттрениран на ImageNet (без топ слоја) Замрзнути конволуционали слојеви у првој фази Fine-tuning последњих Dense блокова Head: GAP + Dense(512) + BN + Dropout(0.5) + Dense(256) + BN + Dropout(0.3) + Softmax(2) Adam ($\text{lr } 10^{-4} \rightarrow 10^{-5}$), CCE loss
ResNet50	<ul style="list-style-type: none"> Преттрениран на ImageNet (без топ слоја) Fine-tuning последњих резидуалних блокова Идентичан класификацијски head као код DenseNet-а Adam ($\text{lr } 10^{-4} \rightarrow 10^{-5}$), CCE loss
EfficientNetB0	<ul style="list-style-type: none"> Преттрениран на ImageNet Fine-tuning последњег convolutional блока Стандардни класификациони head (GAP + Dense слојеви + Dropout) Adam ($\text{lr } 10^{-4} \rightarrow 10^{-5}$)
CheXNet (DenseNet121 на ChestX-ray14)	<ul style="list-style-type: none"> Преттрениран на огромном ChestX-ray14 датасету Замрзнуто тело модела, фино подешавање head-а Head адаптиран за 2 класе (NORMAL, PNEUMONIA) Adam (низак lr), CCE loss
Baseline CNN	<ul style="list-style-type: none"> Два конволуциона блока: Conv(32) + ReLU + MaxPool, Conv(64) + ReLU + MaxPool Flatten + Dense(128) + ReLU + Dropout(0.5) Output: Dense(2) + Softmax Adam optimizer, learning rate 10^{-4}

тенденцију да често класификује нормалне снимке као пнеумонију. F1-score за PNEUMONIA је висок (0.8588), што значи да модел добро детектује болест чак и при различитим односима TP/FP. Ови резултати су у складу са оним што је приказано у CheXNet раду, где DenseNet121 показује најбољи однос дубине, параметара и способности генерализације на рендгенским снимцима [5]. Такође, резултат је у оквирима типичних вредности за DenseNet у задацима двокласне детекције пнеумоније, где AUC углавном пада у распону 0.85–0.92 [3].

B. ResNet50

ResNet50 постиже укупну тачност од 73.9%, али са веома неуравнотеженим перформансама по класама: Recall за PNEUMONIA износи 0.9949, али Recall за NORMAL је само 0.3120. Ово значи да модел готово увек препознаје пнеумонију, али при томе превише често погрешно класификује нормалне снимке као болест. ROC крива показује AUC 0.93, што је високо и указује да

модел има добру укупну дискриминативну способност. Овакво понашање је типично за ResNet50 у случајевима када је скуп података неуравнотежен или када се transfer-learning не стабилизује довољно током fine-tuning-а, што је описано у раду „Pneumonia Detection Using Transfer Learning“, где су ResNet модели показивали осцилације у Recall-у за NORMAL класу [16].

C. EfficientNetB0

EfficientNetB0 показује најслабије резултате међу свим тестираним моделима. Укупна тачност је 59.6%, F1-score за NORMAL је само 0.3262, а AUC износи 0.6254, што је нешто боље од случајног класификатора. Овако слаб учинак EfficientNet-а у овом случају може бити последица неколико фактора: мала величина скupa података, мањи моделски капацитет у односу на DenseNet/ResNet и висока осетљивост на иницијализацију и хиперпараметре fine-tuning-а. Иако је EfficientNetB0 у литератури углавном показивао добре резултате на

већим скуповима, рад „A Review on Detection of Pneumonia...“ показује да EfficientNet понекад има проблеме са стабилношћу на малим скуповима података [3].

D. CheXNet

CheXNet у овој имплементацији постиже 76.4% тачности и AUC од 0.86. Иако би се очекивало да CheXNet буде најбољи модел, јер је преттрениран на огромном ChestX-ray14 скупу, његов резултат није доминантан у односу на DenseNet121. Recall за PNEUMONIA је добар (0.9667), али Recall за NORMAL остаје низак (0.4274), што указује да се доменска разлика између ChestX-ray14 (одрасли пацијенти) и Kaggle датасета (деца старости 1–5 година) одразила на перформансе. Ово је у складу са закључком CheXNet рада, у коме аутори истичу да transfer learning из једне популације у другу може довести до деградације резултата ако се стил снимања значајно разликује.

E. Основни CNN

Основни CNN модел, обучаван од нуле, постиже запажено добре резултате. Укупна тачност је 74.5%, AUC је 0.94, Recall за PNEUMONIA је 0.9923, а Recall за NORMAL је 0.3333. Иако је још увек присутан исти дисбаланс као код ResNet50, тј. висок осећај за PNEUMONIA, а слаб за NORMAL, изненађује да baseline архитектура надмашује ResNet50 у AUC-у и постиже F1-score за PNEUMONIA од 0.8296. Ово је занимљиво, јер показује да модел који се тренира од нуле може бити конкурентан transfer learning приступима на доменски уском и релативно малом скупу података [3].

Литература

- [1] World Health Organization, Pneumonia in children, <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- [2] Wikipedia, Epidemiology of pneumonia https://en.wikipedia.org/wiki/Epidemiology_of_pneumonia
- [3] Alapat, Daniel Joseph, Malavika Venu Menon, and Sharmila Ashok. "A review on detection of pneumonia in chest X-ray images using neural networks." *Journal of biomedical physics and engineering* (2022).
- [4] LeCun, Yann, and Yoshua Bengio. Convolutional networks for images, speech, and time series." *The handbook of brain theory and neural networks* (1998).
- [5] Rajpurkar, Pranav, et al. Čhernet: Radiologist-level pneumonia detection on chest x-rays with deep learning."(2017).
- [6] Pan, Sinno Jialin, et al. Domain adaptation via transfer component analysis." *IEEE transactions on neural networks* 22.2 (2010).
- [7] Huang, Gao, et al. "Multi-scale dense convolutional networks for efficient prediction."(2017).
- [8] He, Kaiming, et al. Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [9] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *International conference on machine learning*. 2019.
- [10] Srivastava, Nitish, et al. Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 (2014).
- [11] Goodfellow, Ian J., et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).
- [12] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." *International conference on machine learning*. PMLR, 2017.
- [13] Frid-Adar, Maayan, et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." *Neurocomputing* 321 (2018).
- [14] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." *Information processing and management* (2009).
- [15] Hanley, James A., and Barbara J. McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology* 143.1 (1982).
- [16] Patel, Maahi, et al. "Pneumonia detection using transfer learning." *International Journal of Engineering Research and Technology (IJERT)* 10.10 (2021).
- [17] Chest X-Ray Images (Pneumonia), <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [18] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. pmlr, 2015.
- [19] Abueed, Mohammed AM, et al. "Pneumonia Detection Using Transfer Learning: A Systematic Literature Review."
- [20] Jain, Rachna, et al. "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning." *Measurement* 165 (2020).
- [21] Adam, Kingma DP Ba J. "A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* 1412.6 (2014).
- [22] Chauhan, Tavishee, Hemant Palivela, and Sarveshmani Tiwari. "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging." *International Journal of Information Management Data Insights* (2021).