

Анализа рендгенских снимака плућа применом дубоких конволуционих и трансформер модела

Analysis of Chest X-ray Images Using Deep Convolutional and Transformer Models

Катарина Радивојевић, Факултет техничких наука, Нови Сад

Студијски програм – ЕНЕРГЕТИКА,
ЕЛЕКТРОНИКА И ТЕЛЕКОМУНИКАЦИЈЕ

Кратак садржај – Овај рад представља примену дубоких конволуционих и трансформер модела за аутоматску класификацију рендгенских снимака грудног коша ради откривања рака плућа. Предложен је каскадни систем који у првом кораку раздваја здраве од патолошких снимака, а у другом врши даљу поделу патолошких случајева на онколошке и неонколошке. Модели су фино подешени на скупу од 20000 снимака различитог порекла, уз примену техника аугментације и уравнотежења класа. Резултати показују високу осетљивост у препознавању болесних случајева и поуздано разликовање онколошких промена, што указује на могућност примене система као помоћног алата у радиолошкој дијагностици.

Кључне речи: : рендгенски снимци, дубоко учење, конволуционе неуронске мреже, трансформери, класификација

Abstract – This paper presents the application of deep convolutional and transformer models for automated classification of chest X-ray images aimed at lung cancer detection. A cascaded system is proposed, where the first stage separates healthy from pathological images, and the second stage further divides pathological cases into oncological and non-oncological. The models were fine-tuned on a dataset of 20,000 images of various origins, using augmentation and class balancing techniques. The results show high sensitivity in detecting diseased cases and reliable differentiation of oncological changes, indicating the potential of the system as an assistive tool in radiological diagnostics.

Keywords: chest X-ray, deep learning, convolutional neural networks, transformers, classification

НАПОМЕНА: Овај рад проистекао је из мастер рада чији ментор је био др Владимир Петровић, ред. проф.

1. УВОД

Рак плућа представља један од најчешћих и најсмртоноснијих облика малигнитета у свету. Иако

рано откривање значајно повећава шансе за успешно лечење, велики број случајева се и даље дијагностикује у узнапредовалој фази болести. Рендгенски снимак грудног коша остаје најдоступнији и најчешће коришћени дијагностички алат, али његова интерпретација у великој мери зависи од искуства радиолога и често је ограничена сличностима између различитих патологија.

Развој дубоких неуронских мрежа омогућио је примену аутоматизованих метода у анализи радиолошких снимака. Конволуционе мреже су се показале изузетно ефикасним у препознавању локалних структура, док трансформер модели, засновани на механизму пажње, омогућавају глобалније разумевање слике.

Циљ овог рада је развој и евалуација каскадног система дубоког учења који прво издваја потенцијално патолошке снимке, а затим врши детаљнију класификацију на онколошке и неонколошке. Оваквим приступом смањује се број пропуштених позитивних случајева и обезбеђује већа поузданост система у реалним клиничким условима.

2. ОПИС СКУПА ПОДАТАКА

За потребе овог истраживања формиран је обједињен скуп података који садржи 20000 рендгенских снимака грудног коша, прикупљених из јавних извора (NIH ChestX-ray14, CheXpert, VinBigData) и приватног скупа компаније Visaris. Снимци су одабрани насумичним избором из сва четири скупа, при чему је примењен критеријум повећане осетљивости на присуство онколошких промена, како би модел био обучен на већем броју клинички релевантних случајева.

Скуп је прилагођен потребама каскадног класификатора, те су снимци разврстани у три категорије: здраве, болесне неонколошке и болесне онколошке. Пре обраде извршена је анонимизација и филтрација нечитљивих снимака, као и конверзија из DICOM у PNG формат ради лакше употребе и компатибилности са библиотекама за дубоко учење.

Да би се модел усмерио искључиво на релевантне делове слике, примењен је сопствени алгоритам за издвајање региона плућног поља, заснован на

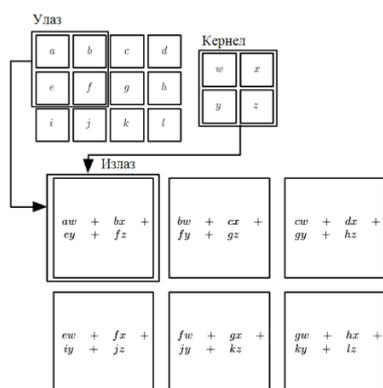
сегментацији. На слици 1 приказан је пример рендгенског снимка пре и после примене овог алгоритма, где се види да се након издвајања задржавају само региони од дијагностичког значаја, док се елиминишу делови позадине и анатомије изван плућног поља.



Слика 1. Слика пре (лево) и слика након (десно) примене алгоритма за издвајање региона плућног поља

3. ТЕОРИЈСКЕ ОСНОВЕ МОДЕЛА

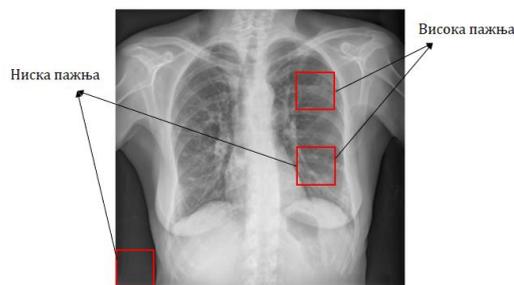
Конволуционе неуронске мреже (CNN) представљају основу већине савремених модела за анализу медицинских снимака. Њихова кључна предност лежи у способности да препознају локалне карактеристике на сликама као што су ивице, сенке и текстуре. Применом конволуционих филтера који пролазе кроз улазну слику. Сваки филтер (језгро) учи да издвоји специфичан образац, док се у дубљим слојевима граде све апстрактније репрезентације, што омогућава моделу да разликује нормалне и патолошке регионе. На слици 2 приказан је принцип рада конволуционог слоја, где се језгро (кернел) креће преко улазне слике и формира излазну мапу карактеристика (*feature map*) на основу прорачуна пондерисаних вредности пиксела [1].



Слика 2. Принцип рада конволуционог слоја [1]

Супротно томе, трансформерске архитектуре (Vision Transformers – ViT) користе механизам „самопажње“ (*self-attention*), који моделу омогућава да истовремено анализира целу слику и одреди које области су најрелевантније за доношење одлуке. Пажња се рачуна између мањих делова слике тј. печева (*patches*), који представљају сегменте слике трансформисане у векторе високих димензија [2]. Скларним производом тих вектора модел придаје већи значај кључним

регионима унутар плућа, као што је илустровано на слици 3, док мање релевантне области добијају мању тежину.



Слика 3. Визуелизација механизма пажње у трансформерској архитектури

4. МЕТОДОЛОШКИ ПРИСТУП

Сви експерименти спроведени су у Google Colab окружењу, уз приступ графичким процесорима (GPU) преко CUDA архитектуре. Истраживање је реализовано у програмском језику Python 3, уз примену библиотеке PyTorch за дефинисање, обуку и евалуацију неуронских мрежа и трансформер модела.

4.1. Избор модела дубоког учења

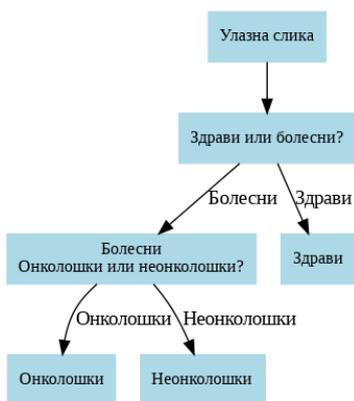
У овом истраживању су од CNN модела испробани ResNet-50 и DenseNet-121, који су се раније показали као изузетно ефикасни у анализи медицинских снимака. ResNet (Residual Network) користи резидуалне везе које омогућавају стабилнији процес обуке и избегавање деградације перформанси код дубљих мрежа. DenseNet (Densely Connected Network) уводи густе повезаности између слојева, што подстиче реупотребу карактеристика и побољшава пропацију градијената [3].

Од трансформерских модела испитани су DeiT (Data-efficient Image Transformer) и Swin Transformer. DeiT је модел развијен од стране компаније Meta (Facebook AI) који представља варијанту ViT-а оптимизовану за обуку на мањим скуповима података, што га чини погодним за ово истраживање. Swin Transformer (Shifted Window Transformer) дели слику на мање, преклапајуће „прозоре“ (*windows*), унутар којих се пажња рачуна локално. Положај прозора се затим помера (*shifted windows*), чиме се омогућава размена информација између суседних региона и постепено гради хијерархијска репрезентација слике [4]. Та особина омогућава прецизније препознавање дисперзних патолошких промена које често карактеришу рак плућа.

4.2. Каскадни приступ класификацији

Да би се повећала осетљивост система и смањено број лажно негативних резултата, примењен је каскадни приступ класификацији, чији се дијаграм може видети на слици 4. Систем се састоји из два модела. Модел А прво класификује рендгенске снимке на здраве и болесне, након чега модел Б анализира само снимке означене као болесне и врши додатну поделу на онколошке и неонколошке. Оваква организација

омогућава постепено сужење анализе и бољу специјализацију сваког модела за свој ниво задатка.



Слика 4. Дијаграм каскадног система класификације

4.3. Припрема података

Формирани скуп од 20000 рендгенских снимака грудног коша подељен је на тренинг (80%), валидациони (10%) и тест (10%) подскуп, уз стратификовану поделу ради очувања односа класа. Пре учитавања у модел, све слике су скалиране на резолуцију 384×384 пиксела, која је одабрана као оптимална за истовремено тестирање конволуционих и трансформерских архитектура. Приликом креирања DataLoader-а примењене су технике аугментације као што су хоризонтално пресликавање, благо прилагођавање контраста и осветљености и додавање Гаусовог шума, док је за патолошке снимке, приликом обуке модела Б, примењена и афина трансформација са смицањем (*shear*).

За уравнотежење класа током обуке коришћен је WeightedRandomSampler, који обезбеђује равномерну заступљеност примера у сваком *batch*-у. Примери снимака пре и након примене ових трансформација приказани су на слици 5, где се може уочити ефекат повећања разноврсности података.



Слика 5. Примери рендгенских снимака пре (горе) и након (доле) трансформације и аугментације

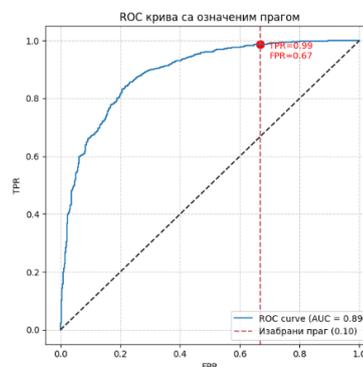
4.4. Подешавање параметара и процес обуке

Сви модели су били претходно тренирани на ImageNet скупу, а затим фино подешени (*fine-tuned*) на скупу од описаном у претходном одељку. При томе је излазни класификациони слој сваког модела замењен новим бинарним слојем, док су претходни слојеви служили као екстрактори карактеристика, што је омогућило

бржу конвергенцију и бољу генерализацију на медицинске слике. За обуку је коришћен AdamW оптимизатор са *weight decay* регуларизацијом и величином *batch*-а 32. Стопа учења је прилагођавана појединачно за сваку архитектуру, а примењен је и *scheduler* који је постепено умањивао њену вредност ради стабилније конвергенције. Функција губитка дефинисана је као *weighted cross-entropy*, са већом тежином за класу „болесни“ у моделу А, односно „онколошки“ у моделу Б, како би се повећала осетљивост на позитивне случајеве. За спречавање пренавикавања примењене су технике *dropout*, *weight decay* и рано заустављање (*early stopping*). Најбољи модели су изабрани према највишој вредности AUC показатеља на валидационом скупу.

5. РЕЗУЛТАТИ И АНАЛИЗА

Након спроведених експеримената, најбоље резултате у првом кораку (модел А) постигао је DenseNet-121, док је у другом кораку (модел Б) најуспешнији био Swin Transformer. Перформансе су оцењене на тест скупу помоћу стандардних метрика: тачност (*accuracy*), прецизност (*precision*), осетљивост (*recall*), F1-скор и AUC. Код модела А посебан акценат стављен је на избор прага класификације ради повећања осетљивости према класи „болесни“. На слици 6 приказана је ROC крива модела А са изабраним прагом класификације (0.1), којим је постигнут компромис између високе осетљивости и прихватљиве прецизности, док је на слици 7 приказана матрица конфузије за модел А. За модел Б, аналогни резултати дати су на сликама 8 и 9, где је уочљиво боље одвајање класа.

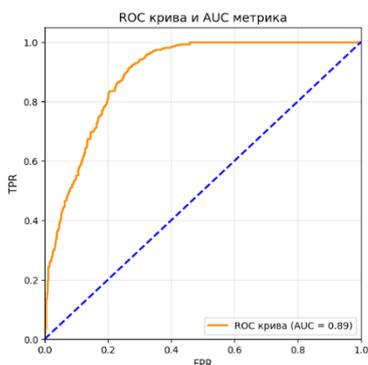


Слика 6. ROC крива и AUC вредност за модел А са одабраним прагом

Матрица конфузије за тест податке

	Здрави	Болесни
Здрави	223	450
Болесни	16	1312
	Здрави	Болесни

Слика 7. Матрица конфузије за модел А (DenseNet-121)



Слика 8. ROC крива и AUC вредност за модел Б



Слика 9. Матрица конфузије за модел Б (Swin Transformer)

Резултати приказани у Табелама 1 и 2 показују перформансе оба модела редом. Модел А је након примене прага постигао укупну тачност од 0.77, уз изразито високу осетљивост за класу „болесни“ (0.99). То указује да модел успешно препознаје готово све патолошке снимке, што је кључно за овај корак. Иако је осетљивост за класу „здрави“ нижа (0.33), овај компромис је оправдан у медицинској дијагностици где је приоритет минимизација лажно негативних случајева. Поред тога, модел је остварио и AUC вредност од 0.89, што потврђује стабилну дискриминативну моћ између класа.

Модел Б је остварио је укупну тачност од 0.82, са уравнотеженим F1 резултатима (0.82 и 0.81). Висока осетљивост за онколошке случајеве (0.93) указује на способност модела да поуздано открије малигне промене, док је за неонколошке снимке постигнута већа прецизност (0.92), што потврђује добру дискриминацију између типова патолошких промена. И код овог модела, постигнута AUC вредност од 0.89 додатно потврђује његову поузданост и стабилност током класификације.

У целини, комбинација ова два модела показује висок потенцијал за примену у каскадним системима, где се у првом кораку обезбеђује максимална осетљивост, а у другом прецизна класификација онколошких случајева.

Табела 1. Мере успешности модела А

	Болесни	Здрави
Прецизност	0.74	0.93
Осетљивост	0.99	0.33
F1 – мера	0.85	0.49
Тачност	0.77	
AUC	0.89	

Табела 2. Мере успешности модела Б

	Онколошки	Неонколошки
Прецизност	0.74	0.92
Осетљивост	0.93	0.72
F1 – мера	0.82	0.81
Тачност	0.82	
AUC	0.89	

6. ЗАКЉУЧАК

У раду је представљен двостепени модел за класификацију, који комбинује предности конволуционих и трансформер архитектура. DenseNet-121 је постигао високу осетљивост у детекцији снимака болесних пацијената, док је Swin Transformer показао већу прецизност у разликовању онколошких и неонколошких случајева. Даља унапређења могла би се остварити проширењем скупа података и применом напреднијих техника предобrade и аугментације, чиме би се побољшала робусност и генерализација модела. Применом оваквих модела могуће је значајно унапредити процес скрининга и раног откривања онколошких промена у плућима, што представља важан корак ка примени вештачке интелигенције у клиничкој пракси.

7. ЛИТЕРАТУРА

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [2] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” ICLR, 2021.
- [3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *CVPR*, 2017.
- [4] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” ICCV, 2021.

Кратка биографија:



Катарина Радивојевић рођена је у Шапцу 2000. год. Мастер рад на Факултету техничких наука у Новом Саду, из области Енергетика, електроника и телекомуникације – смер обрада сигнала, одбранила је 2025.год.

Контакт:

kradivojevic1907@gmail.com