



Аутоматска детекција и класификација патологија на гастроскопским и колоноскопским снимцима

Automatic detection and classification of pathology in gastroscopic and colonoscopic images

Светлана Крунић, Факултет техничких наука, Нови Сад

Студијски програм – ОБРАДА СИГНАЛА

Кратак садржај – Циљ рада је пројектовање и обука аутоматског система за анализу гастроскопских и колоноскопских снимака у видљивом спектру ради детекције присуства патоморфолошких промена и њихове категоризације. Систем је заснован на савременим методама дубоког учења и способан да самостално, на основу снимака насталих током стандардне дијагностичке ендоскопске процедуре, процени да ли у снимцима постоје патоморфолошке промене, које категорије, и представи резултате у облику вероватноће присуства патологије (абнормалан налаз) и вероватноће различитих категорија болести. Употребни циљ је дијагностичка помоћ лекарима приликом дијагностике.

Кључне речи: класификација, ендоскопске слике, машинско учење

Abstract – The aim of this work is to design and train an automatic system for the analysis of gastroscopic and colonoscopic images in the visible spectrum for the detection of the presence of pathomorphological changes and their categorization. The system is based on modern deep learning methods and is capable of autonomously assessing, based on images obtained during standard diagnostic endoscopic procedures, whether pathomorphological changes are present, to which category they belong, and presenting the results as probabilities of pathology presence (abnormal findings) and probabilities of different disease categories. The practical goal was to provide diagnostic support to physicians during the diagnostic process.

Keywords: classification, endoscopic images, machine learning

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

1. УВОД

Развој метода дубоког учења у последњој деценији значајно је унапредио анализу медицинских слика и отворио могућност за примену аутоматизованих система у клиничкој дијагностици. Једна од области у којој ови системи имају велики потенцијал јесте

гастроентерологија, где ендоскопске процедуре, попут гастроскопије и колоноскопије, представљају основни дијагностички алат за визуелну процену слузнице дигестивног тракта. Иако ендоскопија омогућава рано откривање полипа, упалних процеса и других патолошких стања, анализа снимака је захтевна и подложна субјективним грешкама, што указује на потребу за развојем интелигентних система који би подржали лекаре у процесу дијагнозе, скраћујући време дијагностике, лечења, смањујући шансе могућих компликација и побољшавајући квалитет живота пацијената.

У досадашњим истраживањима класификације ендоскопских снимака, посебно на скуповима података HyperKvasir и KvasirV2, доминирају приступи засновани на дубоким конволутивним неуронским мрежама и њиховим ансамблима. Рад *EndoNet* [6] представља један од напредних приступа, заснован на вишестепеној фузији карактеристика добијених из више CNN архитектура (Inception, Xception и ResNet101), уз примену техника редукције и селекције карактеристика као што су NNMF и mRMR. На основу експеримената на HyperKvasir скупу (10 уравнотежених класа), аутори су постигли укупну тачност од 98%.

У раду енгл. “A triple-pronged approach for ulcerative colitis severity classification using multimodal, meta, and transformer-based learning” [7], предложен је свеобухватан модел за класификацију степена тежине улцерозног колитиса, који комбинује мултимодалне моделе, енгл. few-shot meta learning и енгл. Vision Transformer (ViT) архитектуре. Овај приступ је, применом ансамбла више ViT модела, постигао 93% тачности и F1 скор од 93%, уз висок ниво интерпретабилности постигнут анализом SHAP метода.

Рад енгл. “A New Approach for Gastrointestinal Tract Findings Detection and Classification Deep Learning-Based Hybrid Stacking Ensemble Models” [8] представио је хибридни приступ заснован на енгл. stacking ансамблу дубоких неуронских мрежа за детекцију и класификацију налаза у гастроинтестиналном тракту. Аутори су применили статистички валидиран метод са стратификованом поделом података и свеобухватном евалуацијом уз примену више метрика (ACC, F1, MCC, ROC-AUC), при чему је постигнута тачност од 98% на HyperKvasir скупу.

Ипак, већина наведених радова ослања се на спајање више класа ради побољшања квантитативних резултата, што смањује клиничку применљивост и интерпретабилност.

За разлику од њих, у овом истраживању примењен је модел DINOv2 [2], на бази ендоскопским слика отвореног кода HyperKvasir [1], који омогућава учење репрезентација без надзора и показује већу робусност на варијације у осветљењу и текстури. Истраживање је организовано у два нивоа сложености: трокласна класификација (здро, болесно, проблематично) и проширена класификација на 22 класе анатомских и патолошких структура.

Поред тога, евалуација је извршена и на сликама из другог извора тј. гастроскопским снимцима са Клиничко болничког центра Војводине (КБЦВ), чиме је испитана примењивост модела ван јавних скупова података и модел приближен реалним клиничким условима.

2. БАЗА ПОДАТАКА

За реализацију рада коришћен је база отвореног кода HyperKvasir [1] која садржи више од 110.000 слика, видеа и маски лабелираних/нелабелираних патологија и анатомских структура дигестивног тракта. Из скупа је изабран део са лабелираним сликама у видљивом спектру, а након уклањања недовољно заступљених класа, лат. *ileum* и енгл. *hemorrhoids*, формиран је уравнотежен скуп од 11.619 слика распоређених у 22 класе, Табела 1. Подела података извршена је стратификовано у односу 60:15:25 на скупове за обуку, валидацију и тестирање. Прва фаза представљала је класификацију слика у **здро**, **болесно** и **проблематично** ткиво. Трећа класа је неопходна јер представља „остале” класе које нису ни патологије ни анатомске структуре, већ класе које представљају оцену квалитета припреме пацијента за преглед. Ове оцене су важне и неопходне, јер сам квалитет припреме значајно утиче на квалитет дијагностике. Тј. са техничке стране, у присуству артефаката услед лоше припреме пацијента, откривање и класификација патологије није толико поуздана. У другом кораку класификација се фокусира само на релевантне инстанце унутар сваке групе, чиме се постиже прецизнија детекција специфичних ентитета.

Табела 1: Списак класа и број узорака датим класама

Класа	Број слика
Barretts	41
Barretts-short-segment	53
BBPS-0-1	646
BBPS-2-3	1148
Cecum	1009
Dyed-lifted-polyps	1002
Dyed-resection-margins	989
Esophagitis-a	403
Esophagitis-b-d	260
Impacted-stool	131
Polyps	1000
Polyps-masks	1000
Pylorus	999

Retroflex rectum	391
Retroflex stomach	764
Ulcerative colitis grade-0-1	35
Ulcerative-colitis-grade-1	201
Ulcerative-colitis-grade-1-2	11
Ulcerative-colitis-grade-2	443
Ulcerative-colitis-grade-2-3	28
Ulcerative-colitis-grade-3	133
Z-line	932

3. МЕТОД

За решавање задатка примењен је енгл. transfer learning приступ са DINOv2-Base Vision Transformer (ViT) [2] моделом, који садржи 86 милиона параметара, као основом. Модел је коришћен као екстрактор карактеристика, при чему су финални вектори ембединга на CLS токenu улаз у класификаторску главу. У првом моделу класификатор је обучен за три класе, док је у другом моделу извршено фино подешавање за 22 класе.

DINOv2 је визуелни трансформер (ViT), који уместо класичних конволуција користи механизам енгл. self-attention. Слика се дели на енгл. patch-еве величине 16×16 пиксела, који се линеарно пројектују у вектор димензије 768. Секвенца вектора се пропушта кроз 12 трансформер блокова са по 12 глава пажње и скривеним слојем величине 3072, а CLS токен представља глобалну репрезентацију слике која се прослеђује класификаторској глави.

Класификаторска глава модела за класификацију патологија је имплементирана као енгл. fully connected network са следећом структуром: улазни слој 768 неурона, скривени слој 256 неурона са ReLU активацијом, енгл. Dropout слој [4] вредности 0.5 и излазни слој са бројем неурона једнаким броју класа. Класификациона глава трокласног модела садржала је слој за нормализацију и линеарни слој.

Обрада података за оба модела изведена је применом техника обраде слике, као и аугментације ради побољшања робусности модела. Сlike су стандардизоване према статистикама ImageNet скупа и промењене на резолуцију 224×224 пиксела. Употребљен је енгл. WeightedRandomSampler како би се компензовала неуравнотеженост између класа.

Сви експерименти реализовани су у окружењу енгл. PyTorch, а перформансе оба модела праћене су дискутованим метрикама, што је омогућило детаљну анализу тачности по класама.

Трокласни модел је трениран оптимизатором енгл. AdamW [3] са почетном кораком учења 1×10^{-5} , енгл. weight decay параметром 10^{-4} и функцијом губитка енгл. CrossEntropyLoss. Тренинг је изведен током 20 епоха уз енгл. early stopping стратегију након пет епоха без побољшања F1 скор на валидационом скупу.

Тренинг модела за класификацију патологија трениран је такође оптимизатором енгл. AdamW са почетном стопом учења 3×10^{-5} и енгл. weight decay параметром 5×10^{-3} . Класификатор је трениран током 12 епоха, док је функција губитка и овде била CrossEntropyLoss.

4. ЕВАЛУАЦИОНЕ МЕТРИКЕ

Квалитет развијеног модела за класификацију ендоскопских слика оцењен је коришћењем стандардних метрика за вишекласну класификацију. Све метрике су рачунате по класама, а затим приказане и као тежински просек у односу на број узорака сваке класе.

Тачност представља удео исправно класификованих узорака у односу на укупан број узорака:

$$\text{ТАЧНОСТ} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

где су:

TP – број исправних позитивних класификација (енгл. true positives)

TN – број исправних негативних класификација (енгл. true negatives)

FP – број лажно позитивних класификација (енгл. false positives)

FN – број лажно негативних класификација (енгл. false negatives)

Прецизност мери колико су од свих узорака које је модел означио као позитивне заиста позитивни:

$$\text{ПРЕЦИЗНОСТ} = \frac{TP}{TP+FP} \quad (2)$$

Одзив (енгл. Sensitivity или енгл. True Positive Rate) показује колики део стварно позитивних узорака модел успешно препознаје:

$$\text{ОДЗИВ} = \frac{TP}{TP+FN} \quad (3)$$

F1 скор представља хармонијску средину између прецизности и одзива и користи се као уравнотежена мера перформанси модела:

$$F1_СКОР = \frac{\text{ПРЕЦИЗНОСТ} \times \text{ОДЗИВ}}{\text{ПРЕЦИЗНОСТ} + \text{ОДЗИВ}} \times 2 \quad (4)$$

Конфузиона матрица приказује дистрибуцију предикција модела по класама и омогућава идентификацију међукласних грешака. За вишекласну класификацију матрица димензија $K \times K$ садржи вредности:

C_{ij} – „број слика класе i класификованих као класа j “

5. РЕЗУЛТАТИ

У првој фази, у којој је циљ био разликовање здравог, болесног и проблематичног ткива, модел је постигао високу стабилност и тачност класификације. На тест скупу је остварена укупна тачност од 98% и макро F1 мера од 98%, што указује на високу способност генерализације. Највећи број погрешних класификација јављао се између класа „болесно“ и „проблематично“, што је очекивано због морфолошке сличности и присуства артефаката у појединим снимцима.

У другој фази, извршена је класификација на 22 класе анатомских и патолошких региона. Модел је постигао тачност од 91% и тежински F1 скор од 91% на тест скупу. Највећи број мешања класа дешава се у класама степена улцерозног колитиса, где прелазне класе *ulcerative-colitis-grade-0-1*, *ulcerative-colitis-grade-1-2* и *ulcerative-colitis-grade-2-3* често буду класификоване као „пуне класе“ *ulcerative-colitis-grade-1*, *ulcerative-colitis-grade-2* и *ulcerative-colitis-grade-3*. Ово може бити обашњено тиме што је порекло „прелазних класа“ компромис немогућности консензуса од стране стручњака.

Примена модела спроведена је на снимцима добијеним у сарадњи са Клиничко болничким центром Војводине (КБЦВ), који нису били део скупа за обуку, Сlike 1 и 2. За евалуацију коришћене су четири излабеллиране класе (укупно девет слика). У енгл. *zero-shot* тестирању, трокласни модел је успешно препознао све примере у првој фази са 100% тачношћу, док је у другој фази остварио високу сагласност између предвиђених и стварних класа, нарочито код колитиса и ретрофлексираног желуца, Табела 2. Ови резултати потврђују добру генерализацију система и његову применљивост на податке из различитих извора.



Слика 1: Пример слике класе улцерозни колитис 3 из КБЦВ, класификоване у класу улцерозни колитис 2



Слика 2: Пример слике класе полип из КБЦВ, успешно класификоване

Табела 2: Резултати примене модела на сликама КБЦВ

Истинита класа	Први модел	Други модел
Ulcerative-colitis-grade-3	болесно	Ulcerative-colitis-grade-3
Ulcerative-colitis-grade-3	болесно	Ulcerative-colitis-grade-2
Ulcerative-colitis-grade-2	болесно	Ulcerative-colitis-grade-2
Retroflex-stomach	здрaво	Retroflex-stomach
Retroflex-stomach	здрaво	Retroflex-stomach
Polyps	болесно	Pylorus
Polyps	болесно	Dyed-resection-margins

Polyps/retroflex-stomach	болесно	Retroflex-stomach
Polyps	болесно	Polyps

6. ЗАКЉУЧАК

У овом раду представљен је систем за аутоматску класификацију ендоскопских слика гастроинтестиналног тракта заснован на дотретиравању DINOv2 [2] модела дубоког учења. Постигнути резултати показују да DINOv2 [2] архитектура, иако првобитно развијена за општу визуелну анализу, уз примену енгл. transfer learning-а може бити веома ефикасна и у медицинском домену. У поређењу са класичним конволутивним мрежама, као што су ResNet и EfficientNet, DINOv2 [2] показује већу осетљивост на fine текстуралне и морфолошке разлике, што је посебно значајно код гастроентеролошких снимака где су промене често суптилне. Високе вредности тачности и F1 мере у трокласној класификацији потврђују стабилност модела, док резултати у проширеној класификацији показују обећавајуће резултате на којима има места за даље истраживање. Ипак, упркос ограничењима услед неуравнотежености класа и недовољног броја узорака за поједине категорије, систем је задржао робусност и способност да правилно разликује морфолошки сличне класе. Практична примена на стварним узорцима из КБЦВ додатно потврђује потенцијал овог приступа у клиничкој пракси. Добри zero-shot резултати могу се приписати чињеници да су снимци из обе базе настали на сличним ендоскопским уређајима произвођача медицинске опреме Olympus, али и способности DINOv2 [2] модела да препозна глобалне визуелне структуре, а не само локалне шаблоне. Истраживање је показало да је комбинација енгл. self-supervised визуелних репрезентација и енгл. transfer learning приступа ефикасна стратегија за анализу медицинских слика у условима ограниченог броја означених података. Постигнути резултати, са тачношћу до 98% у трокласној и 91% у вишекласној класификацији, потврђују применљивост предложеног модела у клиничком контексту, при чему систем показује способност прецизног препознавања различитих анатомских региона и патолошких промена, као и отпорност на варијације у осветљењу, углу и контрасту снимака. Планирани наставак истраживања укључује увођење DINOv3 [5] архитектуре, проширење скупа података у сарадњи са Клиником за гастроентерологију и хепатологију и развој енгл. multitask модела који би истовремено вршили класификацију и сегментацију лезија, што би додатно повећало клиничку вредност система. Будући рад биће усмерен на проширење скупа података, како у смислу броја слика, тако и у погледу обухватања већег броја патологија и анатомских обележја које систем треба да класификује. Планира се и примена енгл. explainable AI техника, како би се крајњем кориснику обезбедила већа поузданост и транспарентност нашег производа. Такође, биће разматрана интеграција сегментационих метода које ће омогућити просторну локализацију лезија и/или региона од интереса. Развијени модел

представља корак ка практичној примени система вештачке интелигенције у ендоскопској дијагностици, доприносећи унапређењу квалитета и ефикасности клиничке праксе.

7. ЛИТЕРАТУРА

- [1] Borgli, H., Thambawita, V., Smedsrud, P.H. *et al. HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Sci Data* 7, 283 (2020). <https://doi.org/10.1038/s41597-020-00622-y>
- [2] Oquab, Maxime, et al. "Dinov2: Learning robust visual features without supervision." *arXiv preprint arXiv:2304.07193* (2023).
- [3] Loshchilov, Ilya, and Frank Hutter. "Decoupled weight decay regularization." *arXiv preprint arXiv:1711.05101* (2017).
- [4] Srivastava, Nitish, et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research* 15.1 (2014): 1929–1958.
- [5] Siméoni, Oriane & Vo, Van Huy & Seitzer, Maximilian & Baldassarre, Federico & Oquab, Maxime & Jose, Cijo & Khalidov, Vasil & Szafraniec, Marc & Yi, Seungeun & Ramamonjisoa, Michaël & Massa, Francisco & Haziza, Daniel & Wehrstedt, Luca & Wang, Jianyuan & Darcet, Timothée & Moutakanni, Théo & Sentana, Leonel & Roberts, Claire & Vedaldi, Andrea & Bojanowski, Piotr. (2025). DINOv3. 10.48550/arXiv.2508.10104.

DOI: <https://doi.org/10.24867/34BE11Kronic>

DOI: <https://doi.org/10.24867/34BE11Kronic>

DOI: <https://doi.org/10.24867/34BE11Kronic>

Кратка биографија:



Светлана Крунич рођен је у Врбасу 2001. год. Мастер рад на Факултету техничких наука из области Електротехнике и рачунарства – Обрада сигнала одбранила је 2025.год.
Контакт: skrunic43@gmail.com