

Objedinjene upute za web

Struktura inovacijskog natjecanja

Natjecanje se odvija u dva kruga, pri čemu će u prvom krugu natjecateljima biti dostupni javni podaci i pristup Internetu, a u drugom krugu se očekuje rad na hrvatskim podacima unutar sigurnog računalnog okruženja (vidi poglavlje Sigurno računalno okruženje). Prvi krug nije eliminacijski tj. svi timovi imaju pravo pristupa drugom krugu, a konačni pobjednik određuje se na kraju 2. kruga natjecanja.

Tema natjecanja je riješiti dva prediktivna problema: 1.) na temelju mamografskih snimaka utvrditi potrebu za daljnjom medicinskom obradom te 2.) na mamografskim slikama detektirati i lokalizirati suspektne lezije (vidi poglavlje Labele za 2. zadatak). Dodatno, od timova se očekuje da interpretiraju predložena rješenja, razviju prototip korisničkog sučelja nužnog za korištenje rješenja u kliničkoj praksi i opišu potencijal za komercijalizaciju inovacije.

Na kraju drugog kruga natjecanja, konačni poredak timova utvrdit će se bodovanjem ovih zadataka:

1. Predviđanje potrebe za daljnjom medicinskom obradom ("prediktivni zadatak") - do 30 bodova
2. Detekcija i lokalizacija suspektnih lezija - do 30 bodova
3. Interpretacija ponuđenih rješenja - do 20 bodova
4. Prototip korisničkog sučelja i potencijal za komercijalizaciju inovacije - do 20 bodova

Ocjene za Zadatke 1 i 2 temeljit će se isključivo na prediktivnim metrikama rješenja (vidi poglavlje Evaluacija zadataka), dok će Zadatke 3 i 4 ocjenjivati stručno Povjerenstvo za ocjenu rješenja.

Zadaci i evaluacija rješenja

Za prvi i drugi zadatak potrebno je osigurati ponovljivost svih rezultata, što s finalnom predajom rješenja evaluiraju tehničko i medicinsko povjerenstvo za ocjenu rješenja.

1. zadatak: Predviđanje potrebe za daljnjom medicinskom obradom (do 30 bodova)

Cilj ovog zadatka je predvidjeti potrebu za daljnjom medicinskom obradom slučaja, koji je predstavljen kao 4 mamografske slike. Očekivani ishod je predikcija pripadnosti slučaja jednoj od 4 klase, temeljene na BI-RADS sustavu:

- C0, obuhvaća BI-RADS=0: Slučaj zahtijeva daljnju medicinsku obradu drugom dijagnostičkom metodom

- C1, obuhvaća BI-RADS=1 i BI-RADS=2: Slučaj ne zahtijeva daljnju medicinsku obradu drugom dijagnostičkom metodom
- C2, obuhvaća BI-RADS=3: Slučaj zahtijeva praćenje istom dijagnostičkom metodom
- C3, obuhvaća BI-RADS=4 i BI-RADS=5: Slučaj zahtijeva daljnju medicinsku obradu.

Evaluacija rješenja zadatka 1. temeljiti će se na mjeri višeklasni (macro) Mathew's correlation coefficient (MCC). Ljestvica uspješnosti rezultata za zadatak 1 biti će temeljena na ovoj mjeri.

2. zadatak: Detekcija i lokalizacija suspektnih lezija (do 30 bodova)

Cilj ovog zadatka je detekcija i lokalizacija suspektnih lezija na mamografskim slikama.

Očekivani ishod je predikcija pripadnosti detektirane lezije jednoj od klasa suspektnih lezija:

- Masa
- Suspektna kalcifikacija
- Fokalna asimetrija
- Arhitekturnalna distorzija
- Suspektni limfni čvor
- Ostalo – (asimetrije, globalne asimetrije, retrakcija kože, retrakcija bradavice, zadebljanje kože)

te anotacija detektirane lezije na pojedinoj slici. Zlatni standard anotacije su pravokutnici.

Evaluacija rješenja zadatka 2. temeljiti će se na srednjoj vrijednost triju mjera:

- A. Multi-label klasifikacija lezija na pojedinoj slici - macro-average MCC,
- B. IoU - preko multi-label maski lezija po svakoj slici,
- C. IoU - preko združene maske svih lezija na slici.

Ljestvica uspješnosti rezultata za zadatak 2 biti će temeljena na aritmetičkoj sredini navedenih mjera $(A+B+C)/3$.

3. zadatak: Interpretacija ponuđenih rješenja (do 20 bodova)

Cilj ovog zadatka je napraviti pisani izvještaj s opisom razvoja prediktivnih modela iz prvog i drugog zadatka. Ocjenjivat će se analitičke sposobnosti tima i kvaliteta izvještaja koji mora uključivati sljedeće elemente:

- Analiza podatkovnog skupa i objašnjenje razvoja prediktivnih modela prvog i drugog zadatka. Opis cjelokupnog procesa modeliranja, uključujući predprocesiranje, odabir podataka, uzorkovanja, učenje modela i evaluacije.
- Presentacija i interpretacija svih rezultata, kvalitete rješenja i njene potencijalne primjene u stvarnom svijetu.

Pisani izvještaj ocjenjuje tim organizatora natjecanja prema unaprijed definiranim kriterijima, koji će biti detaljno prezentirani na Meetupu nakon završetka 1. kruga natjecanja. Ljestvica

uspješnosti rezultata za zadatak 3 biti će temeljena na aritmetičkoj sredini ocjena pojedinačnih evaluatora.

4. zadatak: Prototip korisničkog sučelja i potencijal za komercijalizaciju inovacije (do 20 bodova)

Cilj ovog zadatka je proizvesti prototip aplikacije za interaktivnu analizu mamografskih snimaka koja koristi prediktivne modele razvijene u prvom i drugom zadatku. Prototip aplikacije mora sadržavati korisničko sučelje (Web based) koje bi se potencijalno moglo koristiti u zdravstvenom okruženju. Prototip aplikacije se može razvijati i na vlastitom računalnom okruženju koristeći javno dostupne podatke. U sklopu ovog zadatka potrebno je pripremiti video pitch u trajanju do 3 minute kojem će se prezentirati poslovni aspekti rješenja

Prototip aplikacije ocjenjuje tim organizatora natjecanja prema unaprijed definiranim kriterijima, koji će biti detaljno prezentirani na Meetupu nakon završetka 1. kruga natjecanja. Ljestvica uspješnosti rezultata za zadatak 4 biti će temeljena na aritmetičkoj sredini ocjena pojedinačnih evaluatora.

Podaci

Unutar sigurnosnog računalnog okruženja, natjecateljima će biti dostupni anonimizirani mamografski podaci (u prvom krugu radi se o javno dostupnim podacima, a u drugom krugu o podacima hrvatskih pacijentica). Put do podataka nalazi se u "data_path.txt" datoteci. Za svaku mamografiju (tzv. slučaj), dostupne su 4 snimke, po 2 za svaku dojku:

1. Kraniokaudalna projekcija - snimanje dojke odozgo prema dolje
2. Kosa mediolateralna (MLO) projekcija - bočno snimanje dojke pod kutem od cca 45°

Slike su dane bez kompresije u .jpeg formatu varijabilne visine i širine (**H x W**). Javni skup podataka dostupan u za prvom krugu natjecanja sadrži oko 17000 slučajeva, dok hrvatski skup podataka dostupan u za drugom krugu natjecanja sadrži oko 35000 slučajeva. Lezije su anotirane i detekrirane u podskupu ovih podataka: 5000 slučajeva u javnom skupu podataka te 500 slučajeva u hrvatskom skupu podataka.

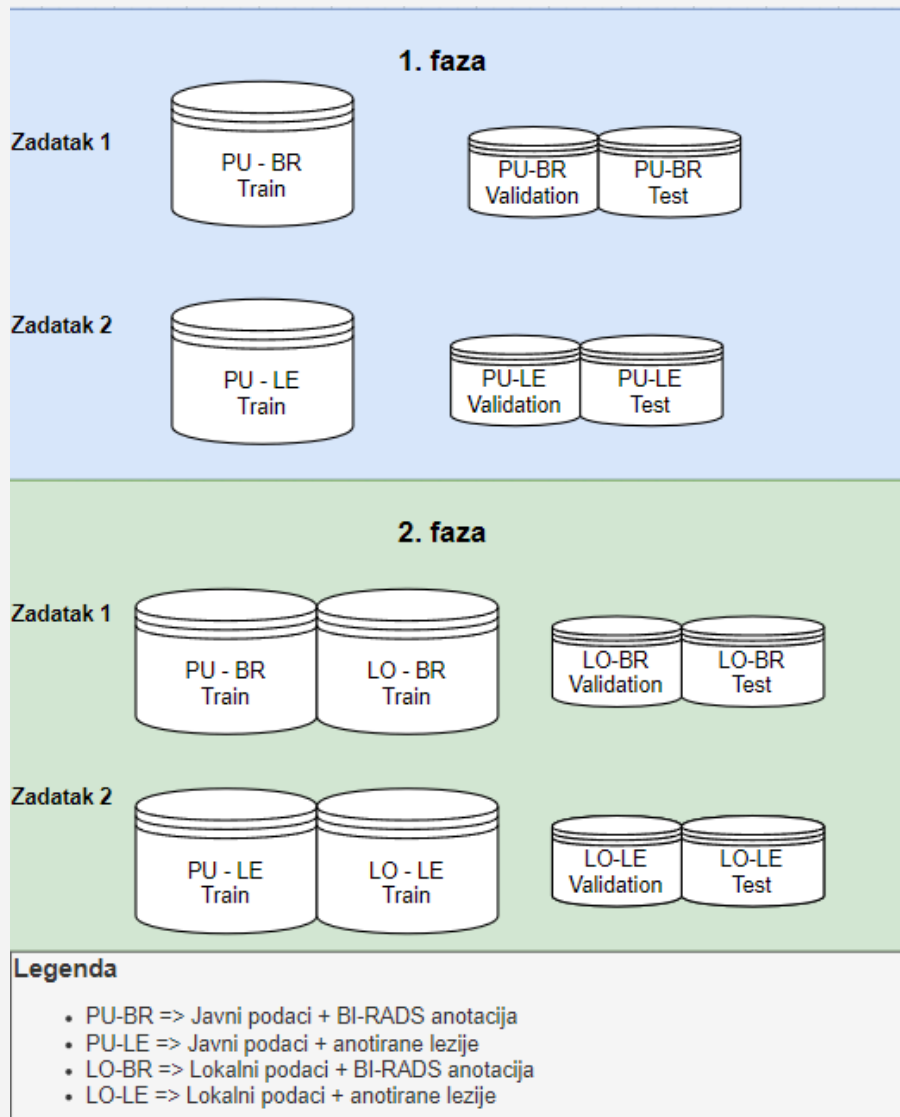
Podatci se nalaze u sljedećim direktorijima:

SRCE: /data/train/

IRB: /home/data/

Napravljen je symbolic link u svaki home direktorij /home/teamXX/data.

Shema podataka po pojedinim fazama natjecanja i zadacima, prikazana je na Slici 1. Važno je napomenuti da je podjela podataka na trening i evaluacijski skup (validation+test) stratificirana, tako da su distribucije klasa slične u trening i evaluacijskom skupu.



Slika 1. Shema podataka po fazama i zadacima natjecanja.

Slanje rezultata

Rješenja zadataka se u oba kruga natjecanja prenose u “submission” mapu unutar računalnog okruženja tima. Rješenja predana po završetku drugog kruga Natjecanja neće se uzimati u obzir.

Predaja rješenja zadataka 1 i 2 sastoji se od kopiranja skripti modela i koda u submission direktorij. Evaluacija ovih rješenja će se odvijati nakon što se aktivira “evaluator” naredba kroz terminal. U tom trenutku organizatori natjecanja preuzimaju dostupni kod i izvršavaju vaš submission.py u okruženju evaluatora. U mapi/direktoriju submission-example dostupan je

primjer. Svako slanje rezultata treba sadržavati rješenja za oba zadatka tj. nije moguće predati rješenje samo za jedan od zadataka.

Na istu lokaciju prenose se rješenja zadataka 3 i 4, koja će se ocjenjivati po završetku drugog kruga Natjecanja.

Za anotaciju testnih primjera ne smije se koristiti ručno označavanje, odnosno ljudsko predviđanje.

Ograničenje na broj evaluacija preliminarnih rješenja

Tijekom natjecanja natjecatelji mogu predavati rješenja zadataka 1 i 2 na preliminarnu evaluaciju. Moguće je predati najviše jedno rješenje dnevno, pri čemu oba zadatka trebaju biti predana.

Ukupno u prvoj fazi natjecanja do najviše 5 rješenja. Za preliminarna rješenja (pokušaje) Organizator će vršiti evaluaciju rješenja na validacijskom podskupu testnog skupa (cca 50%). Konačnu evaluaciju rješenja zadataka 1 i 2 Organizatori će izvršiti na cjelokupnom testnom skupu podataka. Konačna rješenja na testnom skupu podataka moraju biti zaprimljena prije isteka roka za prvi krug Natjecanja.

Za konačnu evaluaciju na testnom skupu koristit će se zadnje poslano rješenje Tima na testnom skupu!

KRUGOVI NATJECANJA

Razvoj rješenja se odvija na udaljenom računalu, pri čemu se u oba kruga natjecanja timovi spajaju na isto radno okruženje. Razlika između prvog i drugog

- Prvi krug natjecanja traje od 5. 2. 2025. - 28. 2. 2025. U prvom krugu natjecatelji postavljaju svoje radno okruženje na udaljenom računalu, pri čemu im je omogućen pristup Internetu i na raspolaganju im je javno dostupan podatkovni set. Natjecatelji imaju slobodu u radnom okruženju postavljati alate i pakete po vlastitom izboru, imajući na umu licencna ograničenja. Nakon završetka prvog kruga natjecanja organizirat će se Meetup koji će predstaviti način bodovanja timova natjecatelja na kraju 2. kruga natjecanja.
- Drugi krug natjecanja počinje nakon Meetupa i traje do- 24. 3. 2025. U drugom krugu će biti onemogućeni pristup Internetu s udaljenog računala i prijenos podataka između lokalnog i udaljenog računala. Natjecatelji rade u sigurnom računalnom okruženju i na raspolaganju im je anonimiziran skup hrvatskih podataka.

Prvi krug AI4Health.Cro inovacijskog natjecanja 2025

Radno okruženje

U prvom krugu natjecatelji postavljaju svoje radno okruženje na udaljenom računalu, pri čemu im je omogućen pristup Internetu i na raspolaganju im je javno dostupan podatkovni set. Predviđamo da će timovi prvi krug natjecanja koristiti za upoznavanje i prilagodbu sigurnog računalnog okruženja, te inicijalni razvoj modela (zadaci 1 i 2). za problem klasifikacije i lokalizacije lezija na mamografskim slikama.

Na udaljeno računalno okruženje korisnici se spajaju s RDP (*remote desktop protocol*) klijentom: korisnici Windows operativnog sustava mogu koristiti Remote Desktop Connection, dok Linux korisnici mogu koristiti Remote Desktop. U sigurnom računalnom okruženju mogu se koristiti isključivo **open-source softveri** i javne biblioteke. Korisnički podaci i upute za spajanje biti će dostavljeni voditeljima timova putem elektronske pošte.

Svakom timu biti će na raspolaganju sigurno računalno okruženje istih karakteristika: LINUX, 6 CPU, 700 GB diskovnog prostora, GPU 20GB.

Podaci i evaluacija

Skup podataka za prvi krug natjecanja je kreiran iz javno dostupnih istraživačkih podataka, te je stratifikacijski podijeljen na trening i test skup (cca 80% / 20%). Natjecateljima je dostupan samo trening skup, dok će testni skup biti korišten za preliminarnu evaluaciju i pružanje povratne informacije o performansama modela.

Tijekom prvog kruga natjecanja, timovi mogu na evaluaciju predati svoje modele tj. rješenja za 1. i 2. zadatak. Korisnici mogu predati najviše jedno rješenje po zadatku dnevno, te maksimalno 5 preliminarnih rješenja za svaki od zadataka tijekom prvog kruga natjecanja. Evaluacija u prvom krugu natjecanja ne utječe na konačnu evaluaciju koja se provodi na kraju drugog kruga natjecanja, već služi kao povratna informacija za daljnji razvoj rješenja. Trenutačni poredak timova baziran na metrikama biti će vidljiv tijekom Natjecanja na web stranici Natjecanja ("Ljestvica rezultata (leaderboard)").

Dodatno, za modele predane na evaluaciju, timovi će dobiti sljedeće povratne informacije o performansama modela:

1. Zadatak
 - a. (macro-average) površinu ispod ROC krivulje
 - b. macro F1
 - c. Imbalanced Multiclass Classification Performance (IMCP) [1]
 - d. za svaku klasu Precision, Recall i MCC
2. Zadatak
 - a. (macro-average) površinu ispod ROC krivulje
 - b. macro F1
 - c. za svaku klasu Precision, Recall i MCC

na validacijskom podskupu evaluacijskog skupa podataka.

Drugi krug AI4Health.Cro inovacijskog natjecanja 2025

Radno okruženje

U drugom krugu inovacijskog natjecanja koriste se sigurna računalna okruženja, tj. računalnim okruženjima koja su timovi koristili u prvom krugu natjecanja se omogućava pristup anonimiziranim hrvatskim podacima, ali im se onemogućava pristup internetu i prijenos podataka između lokalno i udaljenog računalnog okruženja. Timovima ostaju na raspolaganju open-source softveri i javne biblioteke koje su u okruženje instalirali u prvom krugu natjecanja.

EVALUACIJSKE METRIKE

Prvi zadatak

Prvi zadatak kao cilj ima predikciju jedne od 4 klase C0, C1, C2, C3:

BIRADS 0 = C0,

BIRADS 1,2 = C1,

BIRADS 3 = C2,

BIRADS 4,5 = C3

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

MCC je simetrična metrika koja se koristi za binarne klasifikacijske probleme i daje visoke vrijednosti samo ako binarni klasifikator točno predviđa većinu pozitivnih primjera, ali i većinu negativnih primjera.

MCC poprima vrijednosti u intervalu $[-1, +1]$, s ekstremnim vrijednostima -1 i $+1$ (MCC= -1 predstavlja savršeno pogrešnu klasifikaciju, a MCC= $+1$ savršeno točnu klasifikaciju). MCC=0 je očekivana vrijednost za slučajni klasifikator.

Za razliku od MCC, najčešće korištene klasifikacijske mjere točnost i F1 daju pouzdane rezultate samo kada se primjenjuju na uravnotežene skupove podataka, a tipično pogrešne na neuravnoteženim klasifikacijskim problemima.

Gornja formulacija MCC odnosi se na binarni klasifikacijski problem; s obzirom da je naš problem više-klasni mi ćemo za ovaj zadatak koristiti prosječnu vrijednost MCC (macro-average), t.j. računajući srednju vrijednost preko svih klasa:

$$MCC(\text{macro-average}) = (MCC(C0) + MCC(C1) + MCC(C2) + MCC(C3)) / 4$$

Ova mjera koristiti će se kao mjera za rangiranje rješenja u prvom zadatku. Osim ove metrike, za svako preliminarno rješenje izračunavat će se niz metrika koje trebaju poslužiti dubljoj analizi trenutnih performansi prediktivnog modela: (R)osjetljivost, (P)preciznost, F1, AUROC, IMCP (za svaku od klasa i prosječna vrijednost preko svih klasa) rezultati izračuna metrika nalazit će se u **evaluation-results/** direktoriju.

Drugi zadatak

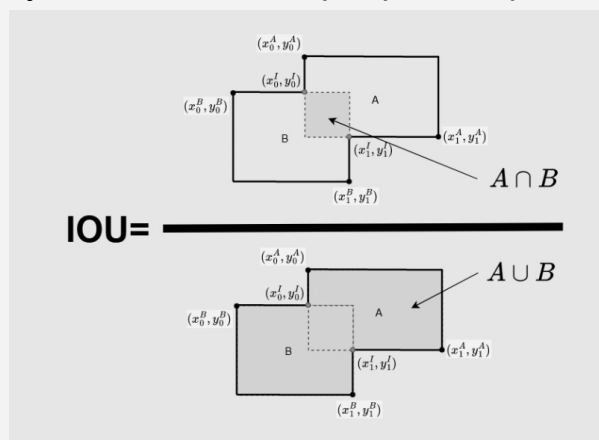
Drugi zadatak sastoji se od dva osnovna pod-zadatka:

2.a) Određivanja tipova lezija prisutnih na slici (tzv. Multi-label klasifikacijski zadatak)

2.b) Semantičke segmentacije po svakoj od 6 tipova lezija, za svaku od slika

Za pod-zadatak 2.a) koristiti ćemo istu mjeru kao i za 1. zadatak, t.j. prosječni MCC preko svih individualnih klasa/labela. Za zadatak 2.b) koristiti ćemo presjek preko unije (IoU - Intersection over Union) mjeru,

izračunavajući presjek i uniju piksela lezija detektiranih modelom i anotiranih od strane radiologa, zasebno za svaku kategoriju lezije. IoU će se za svako preliminarno rješenje izračunati za svaki tip lezije zasebno (IoU_Mass, ..., IoU_Other), te kao prosjek preko svih tipova lezija IoU (macro-average). Npr. A - maska/predikcija za leziju **Mass**, B - oznaka od radiologa lezije **Mass** tada se **IoU(A,B)** računa prema:



Osim ovog prosjeka izračunavat će se i IoU(micro-average) t.j. objedinjavajući predikcije preko svih lezija, odnosno objedinjavajući sve označene lezije na originalnoj slici .

$$\cup(A_{Mass}, A_{Suspicious_Calcification}, \dots, A_{Other}) = C$$

$$\cup(B_{Mass}, B_{Suspicious_Calcification}, \dots, B_{Other}) = D$$

$$\text{IoU} = \frac{C \cap D}{C \cup D}$$

Poput metrika u prvom zadatku, za 2.a) zadatak informat će se natjecatelji sa metrima (R)osjetljivost, (P)preciznost, F1, za svaku od kategorija lezija, kao i prosječna vrijednost preko svih kategorija rezultati izračuna metrika nalazit će se u **evaluation-results/** direktoriju.

Za zadatak 2.b) izračunavati će se IoU za svaku od kategorija lezija, IoU (macro-average), IoU (micro-average).

Ljestvice rezultata

S obzirom na dva zadatka, tokom natjecanja preliminarni rezultati prikazivati će se na 2 ljestvice, zasebno za svaki od problema.

team01, ... team14

Zadatak 1

Macro averages preko BI-RADS kategorija (C0-C3)

RBr	Tim	MCC	F1	R	P	IMCP
1	Team-xx					
2	Team-yy					
3					
...					
13						
14						

Zadatak 2

Macro averages preko kategorija lezija (L0-L5)

RBr	Tim	MCC	F1	R	P	IoU
1	Team-jj					
2	Team-zz					
3					
...					
13						
14						

Napomena: Ovaj izazov predstavlja znanstveni eksperiment, kojim želimo istražiti mogućnosti da se korištenjem različitih izvora podataka, ali i većeg broja različitih pristupa (crowdsourcing) radi dobivanja kompetitivnog modela, odnosno prototipa rješenja potencijalno iskoristivog u medicinskoj praksi. Kao i svaki znanstveni eksperiment, krajnji rezultat nije lako predvidjeti, a tako ni kakvi će se rezultati u smislu performansi na gore definiranim zadacima moći postići.

Stoga će konačni model evaluacije i bodovanja Organizatori definirati nakon završetka prve faze, i dobivanja preliminarnih rezultata od svih učesnika natjecanja. To se ustvari odnosi na definiciju evaluacije i bodovanja za drugi zadatak, odnosno detekciju suspektnih lezija, za koji postoje dvije otežavajuće okolnosti: relativno mali skup javno dostupnih anotiranih slika, te evaluacija na malom broju primjera iz lokalnih izvora.

Literatura

[1] Aguililar-Ruiz & Michalak, 2024, Classification performance assessment for imbalance multiclass data (IMCP) Classification performance assessment for imbalanced multiclass data - https://www.nature.com/articles/s41598-024-61365-z#:~:text=In%20this%20study%2C%20we%20introduce%20the%20Imbalanced%20Multiclass.contrast%20to%20accuracy%20or%20F%20%20%28_beta%29%20-score%29
MCC - <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7>