

Večrazredna klasifikacija obsežnosti COVID-19 okužbe

1 Uvod

V prvem krogu tekmovanja ste se ukvarjali z binarno klasifikacijo obsežnosti COVID-19 okužbe, in sicer med razredoma z omejeno (celokupna CT ocena ≤ 5) in obsežno (celokupna CT ocena ≥ 10) prizadetostjo pljuč. V finalnem krogu tekmovanja klasifikacijski problem razširjamo na 3 razrede, in sicer v učno množico dodajamo CT slike s srednjo prizadetostjo pljuč ($5 < \text{celokupna CT ocena} < 10$).

Vaša naloga je, da v 24 urah sestavite model, ki bo iz CT slik prsnega koša še natančneje napovedal obsežnost COVID-19 okužbe v pljučih.

2 Ozadje

S pomočjo izkušenega radiologa smo na podlagi sistema opisanega v članku označili več kot 1300 CT slik prsnega koša COVID-19 pacientov (nekatero so bile zaradi artefaktov odstranjene). Po tem sistemu je vsakemu pacientu dodeljena celokupna CT ocena, ki predstavlja vsoto ocen po petih pljučnih režnjih. Vsak izmed petih pljučnih režnjev je na podlagi obsega lobarne prizadetosti ocenjen z 0 do 5 točkami. Celokupna CT ocena (vsota po vseh petih režnjih) je tako med 0 in 25.

Na podlagi celokupne CT ocene smo za potrebe prvega kroga tekmovanja slike razdelili v dva razreda. V prvem razredu se nahajajo pacienti z omejeno prizadetostjo pljuč (celokupna CT ocena ≤ 5), v drugem razredu pa pacienti z obsežno prizadetostjo pljuč (celokupna CT ocena ≥ 10). CT slike iz vmesnega intervala, torej pacienti s srednjo prizadetostjo pljuč ($5 < \text{celokupna CT ocena} < 10$), smo v prvem krogu izključili. V finalnem krogu jih dodajamo v učno množico!

3 Podatki

V prvem krogu tekmovanja ste imeli na voljo 355 označenih CT slik iz učne množice (mapa *train*). Informacija o razredu slik se nahaja v datoteki *train.txt*, kjer so slike z omejeno prizadetostjo pljuč označene z indeksom 0, slike z obsežno prizadetostjo pljuč pa z indeksom 1. Te slike ste si prenesli že v prvem krogu (so nespremenjene, 9.4 GB), nahajajo pa se na tej povezavi.

V finalnem krogu v učno množico dodajamo 80 CT slik s srednjo prizadetostjo pljuč ($5 < \text{celokupna CT ocena} < 10$). Vse skupaj imate torej na voljo 355 CT slik z omejeno in obsežno prizadetostjo iz prvega kroga in 80 CT slik s srednjo prizadetostjo iz finalnega kroga. Nove slike (3.0 GB) se nahajajo v mapi *train* na tej povezavi.

4 Navodilo

V finalnem krogu od vas zahtevamo avtomatiziran model, ki bo za neoznačene slike iz finalne testne množice (mapa *final_test* na novi povezavi), napovedal kateremu izmed 3 možnih razredov pripadajo:

- A = omejena prizadetost pljuč (celokupna CT ocena ≤ 5),
- B = srednja prizadetost pljuč ($5 < \text{celokupna CT ocena} < 10$),
- C = obsežna prizadetost pljuč (celokupna CT ocena ≥ 10).

5 Ocenjevanje in oddaja

Vaši modeli bodo ocenjeni na 98 neoznačenih CT slikah iz finalne testne množice. Za vsako izmed CT slik tako podajte črko razreda (A = omejena, B = srednja, C = obsežna prizadetost pljuč). Za razvrstitev ekip se bo uporabila uravnotežena točnost. Primer oddaje v tekstovni datoteki, kjer so imena in napovedi ločene z vejico (vrstni red testnih slik ni pomemben):

```
RIS2021_final_8.nii.gz, A  
RIS2021_final_20.nii.gz, B  
RIS2021_final_58.nii.gz, C  
RIS2021_final_74.nii.gz, A
```

Svoje rezultate **z vso pripadajočo kodo** oddajte v oblak ustvarjen za vašo ekipo. V kolikor boste imeli težave z oblakom, lahko svoje rezultate izjemoma pošljete tudi na ris@dmfa.si (večje datoteke, npr. uteži modela, pa naložite na oblak po svoji izbiri).

V sklopu tekmovanja ne postavljamo omejitev glede uporabljenega programskega jezika, prostodostopnih kod ipd. Edina zahteva je v avtomatiziranem delovanju modela (ročno ocenjevanje slik za namene oddaje ni dovoljeno). Segmentacije pljuč se nahajajo v mapi *segmentations*. Proces predobdelave je prikazan v skripti *preprocessing.py*.

Napovedi v .txt datoteki in **vso pripadajočo kodo** oddajte do nedelje 9.5.2021 do 9:00.