

Binarna klasifikacija obsežnosti COVID-19 okužbe

1 Uvod

Pandemija COVID-19 ni ogrozila le svetovnega zdravja, temveč je močno obremenila in tudi preobremenila zdravstvene sisteme po celotnem svetu, močno spremenila vsakdanje navade in še vedno ogroža svetovno zdravje. Do 2. aprila 2021 je po svetu zbolelo vsaj 130 milijonov ljudi, zaradi COVID-19 pa so umrli že skoraj trije milijoni ljudi. Natančnejše napovedi obsežnosti COVID-19 okužbe v pljučih bi lahko pomagale pri odločitvah o hospitalizaciji in zdravljenju. Ali lahko sestavite model, ki bo iz CT slik prsnega koša čim natančneje napovedal obsežnost COVID-19 okužbe v pljučih?

2 Navodilo

S pomočjo izkušenega radiologa smo na podlagi sistema opisanega v članku označili več kot 700 CT slik prsnega koša COVID-19 pacientov. 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 smo izključili!

Vaša naloga je, da s pomočjo 355 označenih CT slik iz učne množice (mapa *train*) ustvarite model, ki bo čim bolje ločeval med pacienti z omejeno (oznaka 0 v *train.txt*) in obsežno (oznaka 1 v *train.txt*) prizadetostjo pljuč. Vaš model mora biti povsem avtomatiziran, kar pomeni, da na vходу prejme CT sliko na izhodu pa poda verjetnost za pripadnost določenemu razredu (interval $[0, 1]$). Pri tem npr. verjetnost 0.005 predstavlja, da slika z veliko gotovostjo pripada razredu z omejeno prizadetostjo pljuč, verjetnost 0.995 pa, da slika z veliko gotovostjo pripada razredu z obsežno prizadetostjo pljuč.

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*

3 Ocenjevanje in oddaja

Vaši modeli bodo ocenjeni na 151 neoznačenih CT slikah iz testne množice (mapa *test*). Za vsako izmed CT slik tako podajte verjetnost iz zvezne porazdelitve za pripadnost določenemu razredu. Za razvrstitev ekip se bo uporabila metrika AUC. Primer oddaje v tekstovni datoteki, kjer so imena in napovedi ločene z vejico (vrstni red testnih slik ni pomemben):

0006.nii.gz, 0.945

0017.nii.gz, 0.004

0001.nii.gz, 0.543

Svoje rezultate oddajte v oblak ustvarjen za vašo ekipo.