

Krog 1: Klasifikacija slik bakterijskih kolonij

1 Uvod

Pri okužbah zdravniki pogosto vzamejo vzorce kužnine, da ugotovijo, kateri mikroorganizem povzroča okužbo. Eden od načinov identifikacije povzročitelja okužbe je gojenje bakterij na trdih gojiščih, točneje na substratih iz agarja. Vzorec kužnine (npr.: izpljunek - pljučnica, blato - infekcije črevesja, bris rane - infekcije kože) mikrobiologi nanesejo na gojišče. Na površini gojišča se bakterije prično deliti. Ker se potomci ne razpršijo daleč od začetne bakterije, nastane kompaktna skupina celic, ki jo s prostim očesom vidimo kot piko ali madež – to je kolonija. Kolonije različnih bakterijskih vrst se med seboj razlikujejo po barvi, velikosti, obliki roba, prosojnosti in lesku. Mikrobiologi kolonije običajno pregledajo s prostim očesom in uporabijo te lastnosti pri prepoznavanju vrste bakterij.

Vaša prva naloga bo, da razvijete model, ki bo na podlagi barvnih slik (to kar mikrobiolog vidi s prostim očesom) ločeval med različnimi vrstami bakterijskih kolonij.

2 Navodilo

Za prvi krog tekmovanja smo pripravili podatkovno množico 85 RGB slik petrijev z bakterijami, gojenimi na agarju. Slike pripadajo naslednjim 9 razredom:

Ecoli, *Efae*, *Kaer*, *Kpne*, *Paer*, *Saur*, *Sepi*, *Spyo* in *A0*.

Oznaka *A0* pomeni, da je na sliki prikazan samo agar brez ciljne bakterijske kolonije.

Pri delu s podatki upoštevajte naslednje posebnosti:

- nekatere petrijevke so lahko kontaminirane, kar pomeni, da je poleg glavne bakterije lahko prisotna tudi druga, neznana bakterija;
- kontaminacija se lahko pojavi tudi pri slikah razreda *A0*;
- na nekaterih slikah so prisotni artefakti;
- ista petrijevka je lahko slikana večkrat, tudi ob različnih dnevih;
- možno je, da vse oznake v učni množici niso popolnoma pravilne, saj lahko pride do napake pri označevanju.

Vaša naloga je razviti model, ki bo na podlagi učne množice za vsako sliko v testni množici napovedal natanko eno izmed 9 oznak. Gre torej za nalogo večrazredne klasifikacije.

Pri tekmovanju ni omejitev glede izbire programskega jezika, knjižnic, prosto dostopne kode ali javno dostopnih baz podatkov. Ključna zahteva pa je, da mora biti rešitev **popolnoma avtomatizirana**: model mora na vходу sprejeti sliko, na izhodu pa vrniti eno izmed 9 oznak, pri čemer mora delovati **brez dostopa do interneta**. To pomeni, da uporaba oddaljenih API-storitev za klasifikacijo slik (na primer OpenAI API ali podobnih storitev) ni dovoljena. Dovoljena pa je uporaba lokalno nameščenih modelov in orodij, če delujejo brez internetne povezave. Ročno določanje razredov za namene oddaje ni dovoljeno.

Povezava do podatkov: <https://labkey-public.fmf.uni-lj.si/labkey/ris2026/project-begin.view?>

Dostop: Uporabniško ime in geslo sta bila poslana po elektronski pošti.

Pomembno: Uporaba, deljenje ali razširjanje podatkov za namene izven tega tekmovanja ni dovoljeno.

3 Ocenjevanje in oddaja

Vaši modeli bodo ocenjeni na testni množici, ki bo objavljena 22. marca ob 10:00. Rok za oddajo pa bo 23.3. do 12h. Za vsako sliko mora model napovedati en sam razred. Uspešnost napovedi bo ocenjena z metriko: **klasifikacijska točnost** (*accuracy*), tj. z deležem pravilno napovedanih razredov med vsemi slikami v testni množici.

Ob oddaji morate priložiti:

- tekstovno datoteko z napovedmi za testno množico;
- celotno pripadajočo kodo, s katero lahko ponovimo vaše rezultate.

Datoteko z napovedmi in kodo oddajte v oblak, ustvarjen za vašo ekipo.

Datoteka z napovedmi mora biti jasno ločena od izvirne kode in mora biti **obvezno poimenovana z imenom vaše ekipe**.

Format datoteke z napovedmi mora biti naslednji:

IME_SLIKE,OZNAKA

Primer:

```
slikica1.jpg,Ecoli  
...  
slikica26.jpg,A0
```

Pri oddaji pazite, da:

- se ime slike natančno ujema z imenom v testni množici;
- je za vsako sliko podana natanko ena napoved;
- uporabljate izključno dovoljene oznake razredov.