

# Projekt Danica - detekcija zaraženih stanica

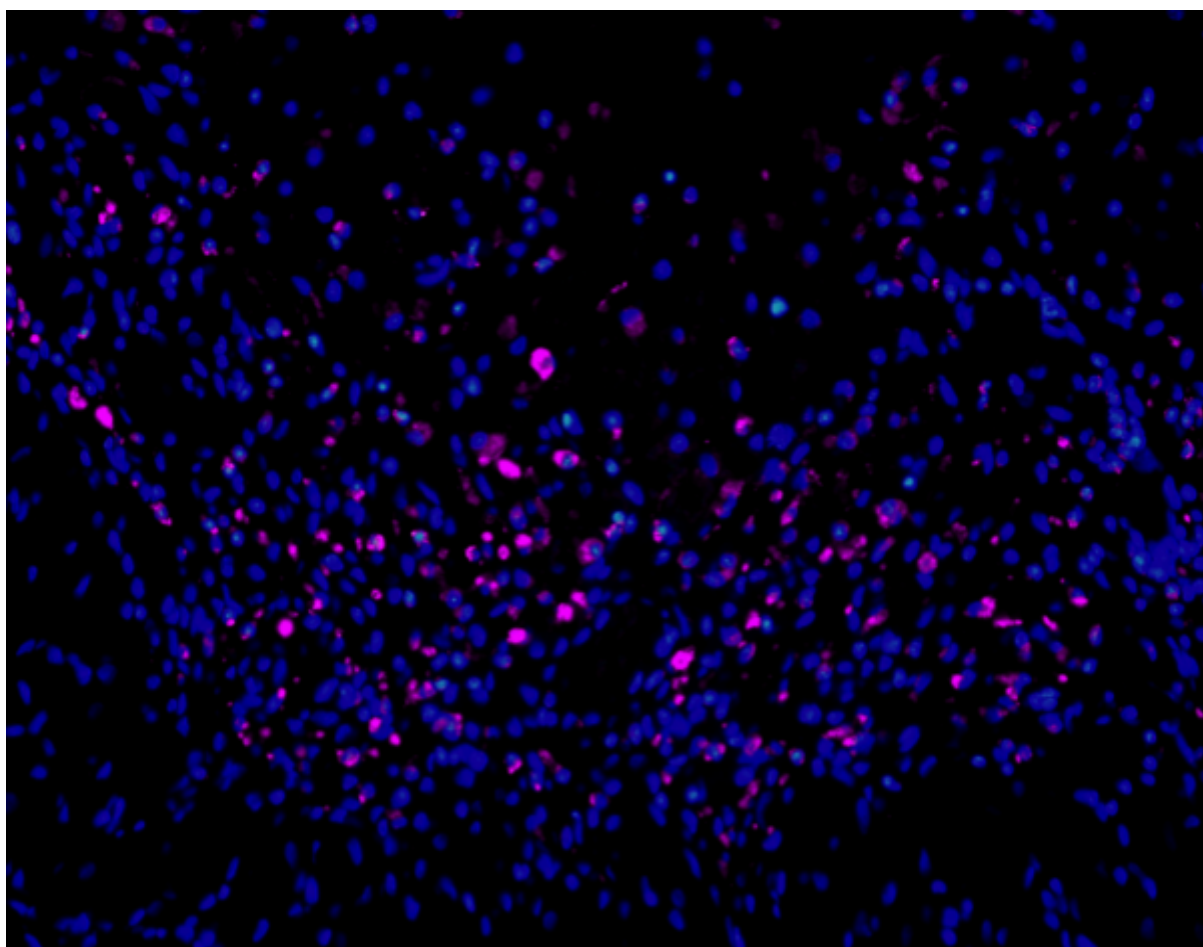
*Al Depope, Marta Han, Ivan Lazarić, Ivan Miošić*

*Travanj 2019*

## Uvod

U istraživanju utjecaja konzumiranja alkohola na rizik od srčanog udara pojavljuje se potreba za određivanjem broja stanica u kojima se odvijaju upalni procesi. Na danom skupu slika prikazane su nakupine stanica čije su jezgre označene plavom bojom, dok je signal koji označava upalni proces roze boje. Naš zadatak je locirati jezgre stanica koje se preklapaju s rozim signalom i pobrojati takvu stanicu kao zaraženu (dakle, ne brojimo rozi signal na poziciji gdje nema stanice te ako ih je više na području jedne stanice, to brojimo kao jednu zaraženu stanicu). Pritom se ne bavimo zaključcima o razlikama između skupina, već nas samo zanima broj zaraženih stanica na danoj slici.

Skup podataka sastoji se od 145 slika dobivenih iz različitih skupina testiranih životinja, a na kojima su izbrojane zaražene stanice i označene lokacije njihovih centara.



*Primjer: Na sljedećoj slici izbrojeno je 177 zaraženih stanica.*

## Generalne ideje i hipoteze istraživanja problema

Iako veličina i oblik stanica variraju, vjerojatno će biti potrebno postaviti ograničenja na veličinu stanice. Naime, vidimo kako postoje mjesta gdje nema jasne granice između nakupine većeg broja stanica pa želimo izbjeći da se takvo područje broji kao jedna stanica kada ih, zapravo, može biti na desetke. Također, u nekim od pristupa rješavanju podijeliti ćemo ove slike na više manjih dijelova, čime ćemo se baviti u nastavku. Pristup nam slijedi iz glavne pretpostavke da je klasifikacija zaraženosti lokalno svojstvo, da zaraženost neke stanice ne ovisi o dovoljno dalekim svojstvima slike.

## Pregled dosadašnjih istraživanja

Problem detekcije stanica često se pojavljuje u raznim istraživanjima u području stanične biologije i medicine te je kao takav kroz godine rješavan brojnim metodama i pristupima. Međutim, pronalazak metode koja bi davala generalno rješenje za širu klasu primjera još uvijek je otvoren problem pa želimo razviti algoritam koji će za ovaj specifični problem davati što točnije rješenje. Većina alata koji rješavaju probleme srodne ovom oslanja se upravo na metode strojnog učenja, a nerijetko su bazirani na neuronskim mrežama. No, korištenje gotovih alata nije se pokazalo pretjerano uspješnim u rješavanju ovog konkretnog problema jer se isti ne sastoji samo od detekcije stanice, već i analize njene okoline. Uz to, nadamo se da će treniranje algoritma na slikama sličnim onima na kojima želimo da on radi rezultirati većom točnošću.

## Materijali, metodologija i plan istraživanja

Dataset Danica sastoji se od ukupno 145 slika, pri čemu je na svakoj slici u prosjeku označeno oko 100 zaraženih jezgara. Obzirom na činjenicu da je originalni skup primjera za treniranje relativno malen, prirodna je ideja podijeliti svaku od slika iz training seta na manje, koje bi idealno sadržavale po jednu jezgru i njenu okolinu, te na takvom, sada bitno većem datasetu, provesti učenje. Jedan način podjele je da prvo parsiramo sliku i nađemo lokacije stanica, čime smo sveli problem na klasifikacijski. Ta metoda je često korištena u radovima slične prirode. Dalje opisujemo našu ideju napada koja se značajno razlikuje od uobičajene, ponajviše u tome što ne reduciramo sliku na stanice na početku.

Slike su dimenzija  $1024 \times 1360$  piksela. Jezgre različitih stanica različitog su oblika i veličina, a ponegdje je čak, što zbog kvalitete slike što zbog bioloških faktora, granica između jezgara nejasna. Stoga će se u projektu konstruirati novi training set na način da će se na slikama iz prvotno određenog training seta na uniformno slučajan način (iz skupa  $[1024/2N, 1024 - 1024/2N] \times [1360/2N, 1360 - 1360/2N]$ ) birati koordinate središta pravokutnika dimenzija  $1024/N \times 1360/N$  piksela. Potom će se kao set za učenje uzeti upravo ti pravokutnici. Broj  $N$  bit će odabran tako da se maksimizira uspješnost na training setu. Kao alternativni pristup određivanju izvedenog dataseta pokušat će se konstruirati pravokutnici slučajne veličine iz intervala  $[100, 1024] \times [100, 1360]$  po gore opisanom pravilu. Dakle, najprije se na slučajan način odredi dimenzija pojedinog pravokutnika, a potom i koordinate njegova središta. Donja ograda na širinu i visinu pravokutnika postoji kako bismo obuhvatili i okolinu većih stanica.

Budući da se ne radi o klasičnom klasifikacijskom problemu u nastavku predložimo sljedeću definiciju ispravno i neispravno pozitivno klasificiranih jezgara: Kažemo da je jezgra stanice J unutar pravokutnika P ispravno pozitivno klasificirana (TP) kao zaražena ako postoji neka označena zaražena jezgra stanice unutar kruga radijusa  $r$  od J u tom pravokutniku. U suprotnom kažemo da je neispravno pozitivno klasificirana (FP). Za  $r$  se uzima radijus najmanje stanice, što prema našoj procjeni iznosi oko 10 piksela. Jezgre stanica koje nisu klasificirane kao zaražene prema gornjoj definiciji, a to jesu, smatramo lažno negativno klasificiranima (FN).

Kao mjeru evaluacije modela koristit će se preciznost, osjetljivost te F1 mjera. Valja imati na umu da u modelu i pozitivnim i negativnim primjerima želimo dati istu težinu. Druge mjere poput IoU (Intersection over Union) ovdje isto nisu pogodne budući da su u originalnom datasetu dani samo podaci o lokaciji središta jezgara zaraženih stanica.

Napomenimo još da je planirano model najprije istrenirati na velikim datasetovima sa sličnom tematikom koji su dostupni na internetu, a tek potom preći na treniranje koristeći vlastiti, gore opisani, training set. To bi se implementiralo na način da "odrežemo" neuronsku mrežu do na zadnjih nekoliko slojeva (postavimo korak u učenju za prethodne slojeve na vrlo malu vrijednost) te takvu neuronsku mrežom tada treniramo s našim podacima.

## Očekivani rezultati predloženog projekta

U sklopu projekta planiramo izvesti model preko kojeg relativno dobro možemo procijeniti broj zaraženih stanica na nekoj slici. Ujedno ćemo analizirati stabilnost generiranja modela s obzirom na odabir testnog skupa. Za većinu izvedenih podataka planiramo imati dostatne vizualizacije.

## Popis literature

<http://jcs.biologists.org/content/126/24/5529>

<https://arxiv.org/pdf/1708.03307.pdf>

<http://www.ece.uah.edu/~dwpan/papers/BHI2017.pdf>

<https://www.annualreviews.org/doi/full/10.1146/annurev-bioeng-071516-044442?intcmp=trendmd>

[https://link.springer.com/content/pdf/10.1007/978-3-642-40763-5\\_51.pdf](https://link.springer.com/content/pdf/10.1007/978-3-642-40763-5_51.pdf)