

Krog 1: Napovedovanje verjetnosti za nastanek raka na dojki

1 Uvod

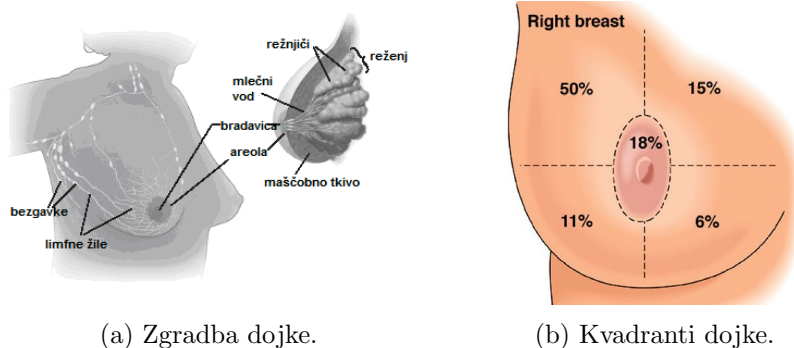
Rak dojke je najpogostejši rak pri ženskah. Letno jih v Sloveniji zbolijo okoli 1400, večina po 50. letu. Iz tega razloga v Sloveniji poteka državni presejalni program za raka dojk (DORA), kjer so ženske med 50. in 69. letom vsaki dve leti vabljeni na presejalno mamografijo, t.j. rentgensko slikanje dojk. Cilj programa je zgodnje odkrivanje raka na dojki. Radiologi poskušajo z mamogramov odkriti, ali se pri pacientki že morda vidijo kakšne sumljive tvorbe in tako omogočiti čimprejšnji začetek zdravljenja. Vaša naloga v prvem krogu tekmovanja pa ne bo odkrivanje raka, ampak napovedovanje verjetnosti za nastanek raka na dojki z normalnih mamogramov. Pacientke, ki so bolj ogrožene (imajo večjo verjetnost, da zbolijo za rakom), bi tako lahko bile slikane pogosteje, tiste z manjšo ogroženostjo pa redkeje.

2 Predstavitev problema

2.1 Rak dojke - medicinsko ozadje

2.1.1 Zgradba dojke

Dojke so parni organ in ležijo nad prsnimi mišicami. So pravzaprav žleze, v katerih se po porodu tvori mleko. Vsaka dojka se deli na 15 do 20 segmentov, imenovanih režnji (latinsko lobul). Režnji vsebujejo številne manjše žlezne režnje, ki so sestavljeni iz mikroskopsko drobnih žleznih mešičkov, v katerih nastaja mleko. Mleko teče iz režnjev skozi mlečne vode (latinsko duktus) v prsno bradavico. Ta se nahaja na sredini temnega dela kože, ki se imenuje kolobar okoli prsne bradavice (latinsko areola). Prostore med žleznimi režnji in mlečnimi vodi izpolnjuje maščobno tkivo. V dojkah se nahajajo tudi limfne žile, po katerih se pretaka bistra tekočina, limfa. Limfne žile vodijo do majhnih okroglih organov, bezgavk. Skupine bezgavk se nahajajo v bližini prsi, v pazduhi, nad ključnico, v prsnem košu za prsnico in v številnih drugih področjih telesa. Bezgavke prestrezajo bakterije, rakave celice in druge škodljive snovi, ki bi lahko bile prisotne v limfnem sistemu. Več informacij na Sliki 1.



Slika 1: Zgradba dojke (a) in kvadranti dojke s pripadajočo verjetnostjo pojavitve raka (b).

2.1.2 Kako nastane rak dojke?

Osnovna značilnost raka dojke kot tudi drugih rakov je nenadzorovana delitev in razrast spremenjenih, rakavih celic. Rak dojke najpogosteje vznikne v izvodilih (latinsko duktus) ali režnjih (latinsko lobul), torej v epitelnih strukturah, zato ga imenujemo tudi karcinom. Če ga od veziva loči bazalna membrana govorimo o neinvazivnem karcinomu (duktalni karcinom in situ – DCIS ali lobularni karcinom in situ – LCIS). Če pa rakave celice prebijejo bazalno membrano in se vraščajo v vezivo dojke, so to invazivni karcinomi. Raki dojke, ki ne vzniknejo iz epitelnih celic in torej niso karcinomi, so redki. To so lahko sarkomi ali maligni limfomi. Ker je delež teh vrst rakov res zelo majhen, sta v klinični praksi rak dojke in karcinom dojke kar sinonima. Če si zamislimo križ s prsno bradavico v središču, je mogoče prostorsko pogostost pojavljanja raka dojke razdeliti na štiri kvadrante (Slika 1). Rak lahko vznikne v katere mkoli kvadrantu, vendar pa je najpogosteje prizadet zgornji zunanji kvadrant, to je del, ki se nahaja najbližje pazduhi, tam pa se nahaja tudi največji del prsnih žlez.

2.1.3 Dejavniki tveganja za razvoj raka dojke

Natančni vzroki nastanka raka dojke niso znani. Znano je, da dotikanje ali stiskanje dojk ne povzroča raka dojke. Rak dojke seveda tudi ni nalezljiv. Raziskave so pokazale, da imajo ženske z določenimi dejavniki tveganja večjo verjetnost, da zbolijo za rakom dojke, kot tiste, ki teh dejavnikov tveganja nimajo. Koristno je poznati dejavnike tveganja, vendar pa je pomembno tudi vedeti, da večina žensk s temi dejavniki tveganja ne zboli za rakom dojke. Dejavniki tveganja so: starost, genetske spremembe (BRCA1, BRCA1), starost ob prvi menstruaciji, starost ob prvem rojstvu, hormonska zdravljenja, prekomerna telesna teža, uživanje alkohola, telesna neaktivnost.

2.1.4 Znaki raka dojke

Znaki raka dojke so različni, največkrat pa se kažejo kot zatrdlina v dojki, sprememba velikosti ali oblike dojke, ugreznjenje prsne bradavice, sprememba kože dojk (pordela, otekla, pomarančasta) in izcedek iz prsne bradavice.

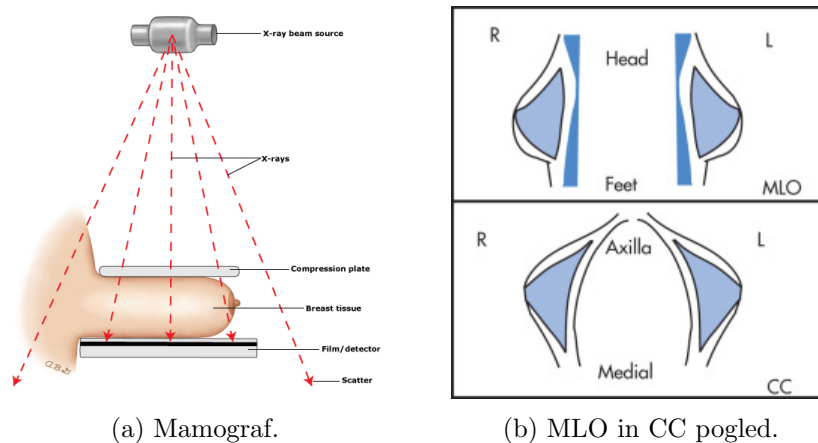
2.2 Motivacija za napovedovanje ogroženosti - družbeno ozadje

Navkljub temu, da vse ženske iz splošne populacije nimajo enake verjetnosti, da bodo zbolele za rakom, so vse vabljeni na presejalno mamografijo z enako frekvenco - na dve leti. Če bi za vsako pacientko posebej vedeli, kakšna je verjetnost, da bo zbolela za rakom v naslednjih letih, bi lahko presejalni program prilagodili. Tiste z večjo verjetnostjo bi slikali z večjo frekvenco kot pa tiste, ki imajo manjšo verjetnost za nastanek raka. S tem bi pri bolj ogroženih pacientkah raka lahko odkrili hitreje.

2.3 Slikanje dojk - fizikalno ozadje

Za preučevanje dojk se uporablja posebna vrsta radiografije, ki jo imenujemo mamografija. Mamografski aparat proizvede snop rentgenskih žarkov in ga izseva na preiskovano območje pacienta - dojko. Snop rentgenskih žarkov se pri prehodu skozi tkivo atenuira. Mesta v dojki, kjer je atomsko število visoko (npr. pektoralna mišica, fibroglandularno tkivo), bodo absorbirala več svetlobe, kot mesta, kjer je atomsko število majhno (npr. maščobno tkivo). Z merjenjem porazdelitve intenzitete izstopne svetlobe dobimo informacijo o anatomski strukturi dojke, kar nam pomaga pri detekciji karakterističnih mas ali mikrokalcinacij. Slika, ki jo dobimo, se imenuje mamogram. Vsako dojka slikamo iz dveh pogledov, kraniokavdalnega (ang. craniocaudal - CC) in mediolateralnega poševnega (ang. mediolateral oblique - MLO), kar pomeni da dobimo 4 slike na pacientko (obstajajo tudi izjeme, a se s tem ne bomo ukvarjali v sklopu tega tekmovanja). Več informacij na Sliki 2.

Če želimo na pridobljenem mamogramu čimbolj ločiti tumorsko od maščobnega in fibroglandularnega tkiva, moramo za to uporabiti fotone relativno nizkih energij. Optimalno območje je med 15 in 25 keV (Slika 3). Posebna značilnost mamografije je tudi kompresija dojke. V primeru, da dojko sploščimo se verjetnost za detekcijo manjših tumorjev poveča (Slika 2).

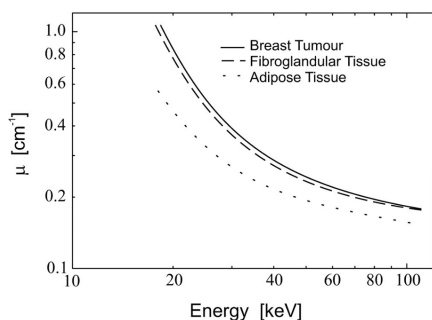


(a) Mamograf.

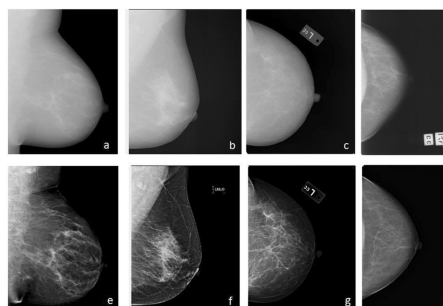
(b) MLO in CC pogled.

Slika 2: (a) Osnovni gradniki mamografa od vrha proti dnu zaporedoma so: izvor rentgenskih žarkov, kompresijska plošča in detektor. (b) Vsaka dojka slikamo iz dveh pogledov: levo-desno predstavlja MLO pogled, zgoraj-spodaj predstavlja CC pogled.

Pred odčitavanjem digitalnih slik je potrebno slike še procesirati. Kontrast med gostim in mehkim tkivom je na surovih slikah nizek, zato te slike niso primerne za vizualno odčitavanje. Vsak proizvajalec je za svoj mamografski aparat razvil tudi svoj algoritem procesiranja slik. Procesirane slike na različnih sistemih so tako različne. Glavni cilj algoritmov je izboljšanje kontrasta med tkivi, kar je razvidno na Sliki 3.



(a) Odvisnost atenuacijskega koeficienta od energije.



(b) Razlika med surovimi in procesiranimi slikami.

Slika 3: (a) Graf prikazuje odvisnost atenuacijskega koeficienta tumorja (polna črta), fibroglandularnega oz. žleznega tkiva (črtasta črta) in maščobnega tkiva (pikčasta črta) v odvisnosti od energije fotonov. Največja razlika med atenuacijskima koeficientoma tumorja in fibroglandularnega tkiva je pri nizkih energijah. (b) Razlika med surovimi (zgoraj) in procesiranimi slikami (spodaj) za različne proizvajalce mamografskih aparatov.

3 Navodilo

Iz presejalnega programa DORA smo za namene prvega kroga tekmovanja zbrali študije 610 pacientk, ki imajo vse normalen mamogram (na mamogramu ni bil zaznan rak). Nekatere bodo v naslednjih 5 letih zbolele za rakom, nekatere pa nikoli (oz. zagotovo ne v naslednjih 5 letih). Vsaka pacientka ima natančno 4 slike, in sicer leva MLO (LMLO), leva CC (LCC), desna MLO (RMLO) in desna CC (RCC). Vse slike so bile zajete na mamografu proizvajalca Hologic. Slike so anonimizirane, procesirane (niso surove) in pretvorjene v PNG16 format. Za pretvorbo v PNG16 format je bila uporabljena knjižnica DCMTK - dcmj2pnm, z nastavitvama +on2 in +Wm.

Za vsako pacientko imate podano informacijo o tem, kdaj v prihodnosti bo zbolela za rakom.

- **Rakave pacientke:** v prihodnjih 5 letih bodo zagotovo zbolele za rakom ($\text{let_do_raka} \leq 5$).
- **Zdrave pacientke:** v prihodnjih 5 letih zagotovo ne bodo zbolele za rakom ($\text{let_do_raka} = 100$).

Vaša naloga je, da s pomočjo označenih slik 610 pacienkt (skupno 2440 slik) ustvarite model, ki bo čim bolj napovedal, ali bo pacientka v naslednjih 5 letih zbolela za rakom ali ne. Z drugimi besedami, ukvarjali se boste z binarno klasifikacijo med razredoma zdravih in rakavih pacientk.

Vaš model mora biti povsem avtomatiziran, kar pomeni, da na vhodu prejme eno ali več slik (vaša odločitev) od izbrane pacientke, na izhodu pa poda verjetnost za pripadnost določenemu razredu (interval $[0,1]$). Pri tem npr. verjetnost 0.005 predstavlja, da pacientka z veliko gotovostjo v prihodnjih 5 letih ne bo zbolela za rakom, verjetnost 0.995 pa, da obstaja zelo velika verjetnost, da pacientka v naslednjih 5 letih bo zbolela za rakom.

V sklopu tekmovanja ne postavljamo omejitev glede uporabljenega programskega jezika, prosto dostopnih kod, javno dostopnih baz ipd. Edina zahteva je v avtomatiziranem delovanju modela (ročno ocenjevanje slik za namene oddaje ni dovoljeno).

Povezava do podatkov: <https://med1.fmf.uni-lj.si/owncloud/index.php/s/aLolUHEFLBADpCc>

Geslo: poslano preko maila prijavljenim ekipam.

Opozorilo: Uporaba in deljenje podatkov za namene izven tega tekmovanja ni dovoljena.

4 Ocenjevanje in oddaja

Vaši modeli bodo ocenjeni na testni množici, ki bo objavljena **27. 3. 2023 ob 15:00**. Za vsako izmed pacientk 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 pacientk ni pomemben):

```
2023278-810, 0.945
2023528-809, 0.004
20233038-809, 0.543
```

Pomembno: za vsako pacientko boste imeli na voljo 4 slike (RCC, RMLO, LCC, LMLO), vendar nas ne zanima, ali vse te slike zares uporabite. Vaša napoved lahko sloni zgolj na eni sliki. Zanima nas zgolj napoved za pacientko. Svoje rezultate oddajte v oblak, ustvarjen za vašo ekipo (povezavo in geslo boste prejeli preko maila).