

Krog 1: Kodiranje patohistoloških izvidov (limfadenektomija in biopsija varovalnih bezgavk)

1 Uvod

Letos se boste soočili z izzivom obdelave in analize tekstovnih podatkov – patohistoloških izvidov, ki nastanejo po kirurških posegih pri bolnikih z melanomom. V Registru raka koderji ročno pregledujejo te izvide in iz njih izluščajo ključne informacije o diagnozi, stadiju bolezni in drugih pomembnih značilnostih. Vaša naloga bo, da razvijete model, ki bo sposoben na avtomatiziran način napovedati vse te spremenljivke iz izvidov.

Ker je ročno kodiranje izvidov zelo zamudno, to tekmovanje predstavlja realno priložnost za prispevek k morebitnim izboljšavam v registru raka. Razvoj učinkovitih avtomatiziranih metod bi lahko v prihodnosti olajšal delo koderjev in povečal hitrost obdelave podatkov.

1.1 Maligni melanom kože

Maligni melanom je šesti najpogostejši rak prebivalcev Slovenije. Bolezen je v Evropi najpogostejša na severu, a tudi v Slovenija se po pogostosti uvršča visoko, na 8. mesto med evropskimi državami. Melanom ima relativno dobro prognozo bolezni, vendar pa se med državami EU najvišja umrljivost beleži na Poljskem in v Sloveniji. Najvišjo incidenco pri nas opazamo v osrednji Sloveniji in na Primorskem. Maligni melanom je en od rakov z najvišjim naraščanjem incidence; v Sloveniji z zadnjih desetih letih starostno standardizirana stopnja narašča za 2% letno.

1.1.1 Patogeneza

Ni popolnoma jasno, zakaj nastane melanom, pomembno vlogo pa ima UV sevanje, svetla polt in genetska predispozicija z mutacijami, kot so BRAF in NRAS. To prav gotovo velja za kožni melanom, ki vznikne iz melanocitov v koži. Poznamo tudi melanom sluznic, ki nastane na lokacijah, ki so skrite pred UV sevanjem, kot so ustna votlina, nosna votlina, analnorektalna regija in uvea očesa. Ta ima drugačne, predvsem genetske dejavnike tveganja, je manj pogost in ima drugačen potek bolezni, zato se ne vodi v Kliničnem registru kožnega melanoma Registra raka in ni vsebovan v tej nalogi, ki vsebuje le patohistološke izvide, povezane z diagnozo kožnega melanoma.

1.1.2 Diagnostični in terapevtski proces

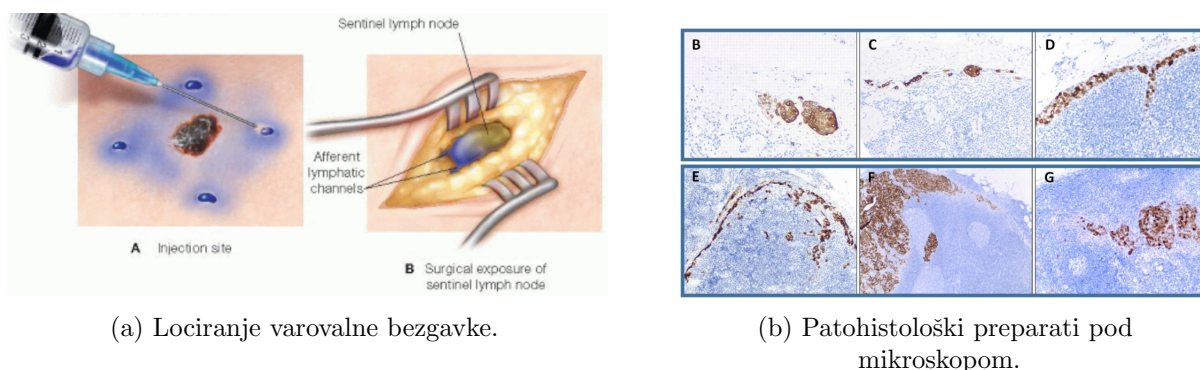
Praviloma so vsi bolniki ob odkritju ali sumu malignega melanoma operirani – kirurgi melanom izrežejo, kar imenujemo ekscizija melanoma. Vzporedno z operacijo oziroma zdravljenjem poteka tudi diagnostični proces, saj so tkiva, odvzeta z namenom zdravljenja, vedno poslana v patohistološki laboratorij na diagnostiko. Pri majhnem melanomu nizkega stadija lahko zadostuje samo ekscizija. Pri večini pa se opravi tudi bolj ali manj obsežna odstranitev bezgavk in njihova patohistološka preiskava.

1.2 Bezgavke pri malignem melanomu kože

Bezgavke so drobne žleze, ki se nahajajo predvsem v vratu, pod pazduhami, v dimljah in okoli notranjih organov in so del limfnega sistema. Filtrirajo limfno tekočino, ki potuje po telesu po limfnih žilah in prenaša celice imunskega sistema. Pomagajo prestrezati rakaste celice, ki so se odcepile od tumorja in po limfnih žilah prišle v bezgavke. V bezgavki lahko zrastejo v nov tumor.

1.2.1 Biopsija varovalne bezgavke

Prva bezgavka, v katero priteka limfna tekočina iz tumorja, se imenuje varovalna bezgavka. Pogosto ta bezgavka prva prestreže tumorske celice, ki se odcepijo od tumorja. Kirurgi pri večini tumorjev te bezgavke (običajno eno, lahko tudi več) odvzamejo in pošljejo na preiskave. Katere so prve drenažne bezgavke, ugotovijo tako, da v brazgotino ob melanomu vbrizgajo radioaktivni označevalec in modro barvilo, ki potujeta iz tumorja po poti drenaže limfne tekočine in se ujameta v drenažni bezgavki, kjer ga zaznajo z uporabo sonde za gama žarke in vizualno (Slika 1a). To bezgavko izrežejo iz tkiva in pregledajo pod mikroskopom za ugotavljanje prisotnosti celic melanoma in njihovih lastnosti (Slika 1b). Če v njej ni rakastih celic, se tumor najverjetneje ni razširil.



Slika 1: Lociranje varovalne bezgavke pred njeno biopsijo (a) in primer patohistoloških preparatov pod mikroskopom (b).

1.2.2 Limfadenektomija

Če se na biopsiji varovalne bezgavke ugotovi prisotnost tumorskih celic, je potrebno odstraniti vse bezgavke na tem območju, opraviti je torej potrebno limfadenektomijo (radikalno disekcijo oz. ekscizijo bezgavk). Pri limfadenektomiji kirurško odstranijo večje število ali vse bezgavke na določenem anatomske območju. Običajno se izvede, kadar je biopsija varovalne bezgavke pozitivna in je potrebno odstraniti preostale bezgavke na območju, ali pa kadar so prisotne metastaze v tipno povečanih bezgavkah. Bolniki ob višjem stadiju bolezni prejmejo tudi ostale oblike zdravljenja (sistemsko zdravljenje ali kemoterapija, obsevanje ali radioterapija in imunsko oz. biološko zdravljenje).

1.3 Kodiranje melanomov v Registru raka

1.3.1 Register raka

Register raka je ustanova, ki deluje znotraj Onkološkega inštituta v Ljubljani in sistematična zbira, shranjuje in analizira podatke o bolnikih z diagnosticiranim rakom z namenom izboljšanje obravnave bolnikov in sistema onkološkega zdravstva. Glavno podatkovno zbirko Registra raka predstavlja populacijski register raka, ki zajema celotno populacijo oseb s stalnim bivališčem v Sloveniji, ki jim je bilo odkrito rakavo obolenje. Register raka neprekinjeno deluje že od leta 1950 in tako upravlja z zbirko podatkov o bolnikih z eno najdaljših časovnih vrst na svetu. Register raka prejme informacije iz zdravstvenih ustanov, ki po zakonski dolžnosti poročajo o vsakem novem primeru rakave bolezni. Register zaposluje koderje, to je zdravstveno osebje, največkrat medicinske sestre, ki opravijo dodatna specializirana izobraževanja. Koderji vso prejeto zdravstveno dokumentacijo preberejo, iz njih izluščijo informacije in jih preko aplikacije Registra raka vnesejo v podatkovno bazo.

1.3.2 Klinični register kožnega melanoma

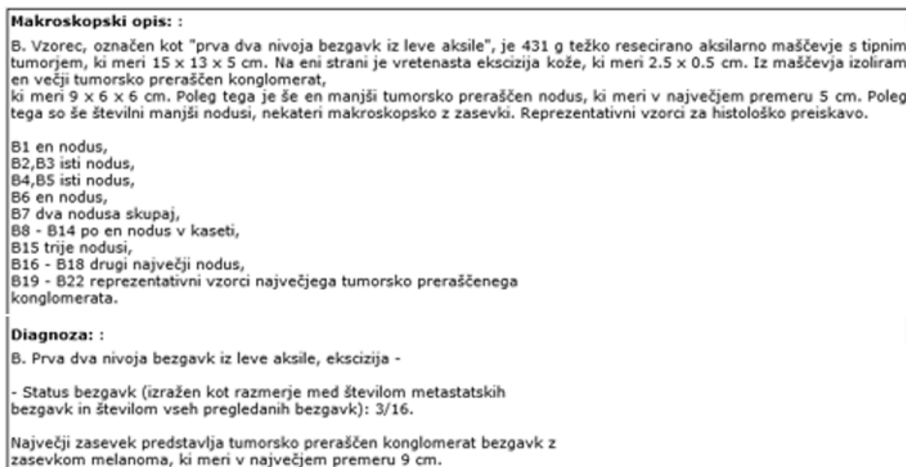
Od leta 2017 Register raka vzpostavlja tudi klinične registre za pet najpogostejših vrst raka, razen nemelanomskega kožnega raka, ki v slovenski populaciji zavzemajo več kot polovico vseh rakov. Za te

rake se beleži razširjen nabor podatkov o diagnostičnem in terapevtskem procesu, ki služijo za oblikovanje kazalnikov o kakovosti zdravstvene oskrbe (npr. čas obravnave bolnika, skladnost obravnave s smernicami, zapleti pri zdravljenju). Enega kliničnih registrov predstavlja tudi Nacionalni klinični register kožnega melanoma. Vanj vnašajo podatke o vseh zbolelih z *invazivnim* in *in situ* kožnim melanomom (C43 in D03 po mednarodni klasifikaciji MKB-10).

1.3.3 Kodiranje malignega melanoma

Pri malignem melanomu predstavljajo osnovni vir podatkov patohistološki izvidi, ki jih Registru pošiljajo vsi slovenski laboratoriji za patologijo, in jih sestavijo zdravniki patologi po mikroskopskem pregledu tkiva. En izvid običajno vsebuje opis več preparatov oziroma delcev tkiva, ki so bili odvzeti skupaj in bili poslani na preiskavo. Kodirati je potrebno pozitivne izvide, torej tiste, v katerih je bila najdena patološka najdba, npr. rakave celice ali zasevek (metastaza), negativni izvidi pa so pomembni za določitev stadija bolezni.

Patološki izvidi so običajno besedilo v nekaj odstavkih (glej primer na Sliki 2). Razdeljen je po posameznih delih, kot so »Klinične diagnoze in podatki«, »Makroskopski opis« in »Diagnoza«. Do neke mere gre za strukturirano oziroma standardizirano besedilo, oblikovano v skladu s slovenskimi Priporočili za obravnavo bolnikov s kožnim melanomom, vendar se izvid nahaja v obliki prosto tvorjenih povedi. Izvidi se lahko razlikujejo med laboratoriji in med patologi znotraj laboratorija. Nekateri podatki lahko manjkajo, kar je eden pokazateljev nižje kakovosti izvida. Zaradi določene mere strukturiranosti izvidov koderjem večine ni težavno najti in določiti. Pri nekaterih informacijah je potrebno besedilo podrobneje analizirati (na primer za določanje največjega premera izmed vseh omenjenih mer zasevkov). Vendar tudi pri strukturirano podanih informacijah težavo predstavlja velika količina informacij, saj je potrebno vsako najdeno informacijo ročno vnesti v aplikacijo. Gre torej za zamudno delo, ki zahteva veliko zbranosti, potrpežljivosti in pomeni tudi visoke stroške zaradi velike količine podatkov in ročnega kodiranja.



Slika 2: Izsek iz patohistološkega izvida, kot ga vidi koder Registra raka v bolnišničnem informacijskem sistemu.

2 Navodilo

Za potrebe prvega kroga tekmovanja smo zbrali 714 tekstovnih izvidov, ki opisujejo kirurške posege, opravljene pri različnih bolnikih v obdobju med letoma 2015 in 2024. Izvide so pripravljali različni patologi. Vsebina izvidov zajema opise dveh vrst kirurških posegov: limfadenektomije in biopsije varovalnih bezgavk.

Vsak izvid je opremljen s sledečimi informacijami:

1. ST_IZVIDA (identifikacija)
2. IZVID (besedilo o opravljenem kirurškem posegu, iz katerega koderji razberejo ostale spremenljivke)
3. PATOLOG (anonimizirana identifikacijska številka patologa, ki je pripravil izvid)
4. LETO (leto priprave izvida)

Koderji iz Registra raka so iz vseh izvidov identificirali 10 spremenljivk. Te spremenljivke, skupaj z njihovim opisom in možnim naborom vrednosti, so naslednje:

1. VRSTA_KRG (vrsta kirurškega posega, ki je bil opravljen pri bolniku)
 - Biopsija varovalne bezgavke
 - Limfadenektomija
2. ST_BEZGAVCNIH_LOZ (število odstranjenih bezgavčnih lož)
 - 1 - 5 (numerična spremenljivka)
3. ST_PREGLEDANIH_BEZG (število histološko pregledanih bezgavk, večje kot je, večja je natančnost določitve stadija bolezni)
 - 1 - 99 (numerična spremenljivka)
4. ST_POZITIVNIH_BEZG (število bezgavk, v katerih so bile najdene tumorske celice)
 - 0 - 82 (numerična spremenljivka)
5. PREMER_MAX_ZASEVKA (največji premer metastatskega zasevka v milimetrih)
 - 0,1 - 99,9 (numerična spremenljivka)
6. POLOZAJ_MAX_ZASEVKA_BEZG (anatomska lokacija največjega metastatskega zasevka v bezgavkah, ki pomaga pri napovednih modelih širjenja bolezni)
 - Drugo
 - Ne
 - Neznano
 - Oboje
 - Parenhimski
 - Subkapsularno
7. EKSTRAKAPS_SIRJENJA (ali se metastatske celice širijo izven kapsule bezgavke)
 - Da
 - Ne
 - Neznano

8. ZRASCENOST_PATO_BEZG (ali so bezgavke zaraščene s tumorskim tkivom)

- Da
- Ne
- Neznano

9. TUMOR_ZUNAJ_BEZG_ZILJA (ali so tumorske celice prisotne zunaj bezgavk v žilah)

- Da
- Ne
- Neznano

10. RADIKALNOST_KRG_BIOPS (ali je bila biopsija ali limfadenektomija izvedena radikalno)

- NEZNANA RADIKALNOST OPERACIJE (OPERACIJA JE BILA)
- R0, NI OSTANKA TUMORJA
- R1, MIKROSKOPSKI OSTANEK TUMORJA
- R2, MAKROSKOPSKI OSTANEK TUMORJA

Izmed vseh izvidov smo naključno izbrali 286 primerov (testna množica) in iz njih odstranili ročno odčitanih 10 spremenljivk. Vaša naloga je, da na podlagi 428 kodiranih izvidov (učna množica) razvijete model, ki bo čim natančneje napovedal manjkajoče spremenljivke v testni množici. Pri tem vam ne zagotavljamo, da so vse ročno odčitane vrednosti spremenljivk iz izvidov pravilne.

V sklopu tekmovanja ne postavljamo omejitev glede uporabljenega programskega jezika, prosto dostopnih kod ali javno dostopnih baz. Edina zahteva je **avtomatiziran model, ki mora delovati brez dostopa do interneta**. To pomeni, da npr. uporaba OpenAI (ali drugih) API klicev za iskanje spremenljivk iz izvidov ni dovoljena. Dovoljeno pa je, na primer, uporabljati lokalne različice prosto dostopnih LLM-jev (DeepSeek, Llama ...). Ročno iskanje spremenljivk iz izvidov za namene oddaje ni dovoljeno.

Povezava do podatkov: <https://labkey-klimt.fmf.uni-lj.si/labkey/ris2025/project-begin.view?>

Dostop: Uporabniško ime in geslo je bilo poslano preko maila.

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

3 Ocenjevanje in oddaja

Vaši modeli bodo ocenjeni na testni množici (286 izvidov), ki bo objavljena **14. 3. 2025 ob 15:00**. Za vsak izvid mora vaš model določiti vrednosti desetih manjkajočih spremenljivk. Uspešnost napovedi bo ocenjena z metriko povprečne točnosti preko vseh desetih spremenljivk.

Svoje napovedi **skupaj s pripadajočo kodo**, s katero lahko poustvarimo vaše rezultate, oddajte v oblak, ustvarjen za vašo ekipo.

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

Primer oddaje (**obvezno dodajte imena stolpcev** - če imena stolpcev ne bodo vključena, bo privzet vrstni red spremenljivk iz učne množice):

```
ST_IZVIDA, VRSTA_KRG, ST_BEZGAVCNIH_LOZ, ..., RADIKALNOST_KRG_BIOPS
121,"Biopsija varovalne bezgavke", "1", ..., "R1, MIKROSKOPSKI OSTANEK TUMORJA"
...
621,"Limfadenektomija", "1", ..., "R0, NI OSTANKA TUMORJA"
```