

Lara Dular<sup>1</sup>, Gregor Brecl Jakob<sup>2</sup>, Lina Savšek<sup>3</sup>, Jožef Magdič<sup>4</sup>, Bojan Rojc<sup>5</sup>, Žiga Špiclin<sup>6</sup>

## Napovedovanje kliničnega poteka multiple skleroze iz magnetnoresonančnih slik: predhodni rezultati raziskave

*The Prediction of Multiple Sclerosis Disease Course from Magnetic Resonance Images: Preliminary Study Results*

### IZVLEČEK

KLJUČNE BESEDE: napovedovanje poteka multiple skleroze, analiza MR-slik, strojno učenje napovednih modelov

**IZHODIŠČA.** Zgodnje prepoznavanje bolnikov z multiplo sklerozo, pri katerih obstaja tveganje za napredovanje bolezni in/ali prehod v sekundarno progresivno multiplo sklerozo, je pomembna neizpolnjena klinična potreba. Namen raziskave je učenje in vrednotenje napovednih modelov napredovanja invalidnosti pri bolnikih z multiplo sklerozo iz MR-slik. **METODE.** Iz T1-poudarjenih slik in slik z izbranim signalom vode smo s samodejno analizo določili prostornino 268 možganskih struktur in prostornino ter število poškodb bele možganovine. S temi meritvami smo učili tri uveljavljene modele razvrščevalnikov, četrti model pa je temeljil na konvolucijski nevronske mreži in se je učil neposredno na sivinskih vrednostih MR-slik. **REZULTATI.** Najboljši rezultat je dosegel četrti model z vrednostjo površine pod krivuljo delovanja sprejemnika 74 %, točnostjo 76 % in občutljivostjo 51 %, model na osnovi razvrščevalnika z naključnimi gozdovi pa je dosegel najvišjo občutljivost 54 %. Navedeni rezultati se skladajo z rezultati v preteklih raziskavah. **RAZPRAVA.** Zmožnost vnaprejšnjega prepoznavanja povečanega tveganja napredovanja invalidnosti neposredno iz MR-slik je obetavna raziskovalna smer. Nadaljnje izboljšave napovednih modelov bi omogočile predčasno ukrepanje, da se lahko prepreči ali odloži napredovanje invalidnosti in/ali prehod v sekundarno progresivno multiplo sklerozo.

<sup>1</sup> Asist. Lara Dular, Laboratorij za slikovne tehnologije, Katedra za biomedicinsko tehniko, Fakulteta za elektrotehniko, Univerza v Ljubljani, Tržaška cesta 25, 1000 Ljubljana

<sup>2</sup> Asist. dr. Gregor Brecl Jakob, dr. med., Klinični oddelek za bolezni živčevja, Nevrološka klinika, Univerzitetni klinični center Ljubljana, Zaloška cesta 2a, 1000 Ljubljana

<sup>3</sup> Asist. dr. Lina Savšek, dr. med., Nevrološki oddelek, Splošna bolnišnica Celje, Oblakova ulica 5, 3000 Celje; Medicinska fakulteta, Univerza v Ljubljani, Vrazov trg 2, 1000 Ljubljana

<sup>4</sup> Asist. Jožef Magdič, dr. med., Klinika za nevrologijo, Univerzitetni klinični center Maribor, Ljubljanska ulica 5, 2000 Maribor

<sup>5</sup> Doc. dr. Bojan Rojc, dr. med., Nevrološka dejavnost, Oddelek za interno medicino, Splošna bolnišnica Izola, Polje 40, 6310 Izola; Oddelek za psihologijo, Fakulteta za matematiko, naravoslovje in informacijske tehnologije, Univerza na Primorskem, Glagoljaška ulica 8, 6000 Koper; Katedra za fizioterapijo, Fakulteta za vede o zdravju, Univerza na Primorskem, Polje 42, 6310 Izola

<sup>6</sup> Izr. prof. dr. Žiga Špiclin, Laboratorij za slikovne tehnologije, Katedra za biomedicinsko tehniko, Fakulteta za elektrotehniko, Univerza v Ljubljani, Tržaška cesta 25, 1000 Ljubljana; ziga.spiclin@fe.uni-lj.si

## ABSTRACT

KEY WORDS: prognosis of multiple sclerosis disease course, MR image analysis, machine learning of prognostic models

**BACKGROUND.** Early identification of multiple sclerosis patients at risk of disease progression and/or progression to secondary progressive multiple sclerosis is an important unmet clinical need. This research aimed to train and evaluate predictive models of disability progression in patients with multiple sclerosis from MR images. **METHODS.** The volumes of 268 brain structures and the volume and number of white matter lesions were determined from T1-enhanced and water-attenuated MR images through the automatic analysis of these images. We used these measurements to train three established classifier models, while the fourth model was based on a convolutional neural network and was trained directly on the gray values of MR images. **RESULTS.** The fourth model achieved the best area under the receiver operating characteristic curve value of 74%, the accuracy of 76% and sensitivity of 51%, and the random forest classifier-based model achieved the highest sensitivity of 54%. The stated results are consistent with the results of previous research. **DISCUSSION.** The ability to identify increased risk of disability progression from MR images is a promising research direction. Further improvements in predictive models would enable early intervention to prevent or delay the progression of disability and/or the transition to secondary progressive multiple sclerosis.

## IZHODIŠČA

Potek multiple skleroze (MS) se precej razlikuje med posameznimi bolniki. Pri večini bolnikov (nad 85 %) je prisotna recidivno-remitentna oblika multiple skleroze (RRMS), za katero so značilna ponavljajoča občasna poslabšanja oz. zagoni boleznin in kasnejše remisije (1). Poznejše (trajno) napredovanje invalidnosti in pojav sekundarno progresivne oblike multiple skleroze (SPMS) veljata za pomemben mejnik v kliničnem razvoju RRMS, katerega značilnost je postopno poslabšanje z malo ali brez zagonov. Oblika SPMS naj bi se v populaciji bolnikov z RRMS razvila v 2–3 % na leto. Kljub razpoložljivosti zdravljenja, ki vpliva na imunsko odzivnost in spreminja potek boleznin med zagoni, pri več kot 80 % bolnikov z RRMS sčasoma bolezen preide v obliko SPMS. Za učinkovito zdravljenje bolnikov, pri katerih obstaja večje tveganje za napredovanje invalidnosti zaradi boleznin in prehod v obliko SPMS, je sposobnost pre-

poznavanja povečanega tveganja za napredovanje boleznin izjemnega pomena.

Za napovedovanje prihodnjega napredovanja boleznin, merjenega z razširjeno lestvico stopnje prizadetosti (Expanded Disability Status Score, EDSS), so v nedavnih raziskavah uporabili razvrščevalnike, ki temeljijo na strojnem učenju (angl. *machine learning*, ML). Branco in sodelavci so za 220 bolnikov napovedovali povečanje EDSS po triletnem obdobju na podlagi odgovorov iz vprašalnikov (angl. *patient reported outcome measures*), pridobljenih v času osnovnega obiska. Preizkusili so več razvrščevalnikov, kot so logistična regresija, naključni gozdovi (angl. *random forests*, RF), naivni Bayesov razvrščevalnik in metoda podpornih vektorjev (angl. *support vector machines*, SVM). Z modelom RF so dosegli najvišjo občutljivost in specifičnost (85 %) (2).

V podobni raziskavi so Pinto in sodelavci na podlagi podatkov (kot so starost, spol, starost ob začetku boleznin, klinični

podatki o zagonih, EDSS in simptomi 187 bolnikov zadnjih petih let) usposobili in primerjali modele ML, učene s podatki, zbranimi v različno dolgih časovnih obdobjih, npr. zadnjih dveh, treh, štirih in pet let. Najboljše rezultate s površino pod krivuljo delovanja sprejemnika (angl. *area under receiver operating curve*, AUROC) 86 %, občutljivostjo 76 % in specifičnostjo 77 % so dosegli z modelom SVM in podatki iz zadnjih dveh let opazovanja (3).

Napredovanje bolezni so napovedovali tudi neposredno iz MR-slik z modeli globokega učenja (angl. *deep learning*, DL). Na naboru podatkov 181 bolnikov so Taloni in sodelavci učili in vrednotili model DL ResNet50 na skladih petih zaporednih rezin T1-poudarjenih slik. Prepoznali so pet koronalnih in pet sagitalnih skladov z vrednostjo AUROC, večjim od 72 oz. 81 % (4). Storelli in sodelavci so usposobili napovedni model DL na T1- in T2-poudarjenih MR-slikah 373 bolnikov, da bi napovedali povečanje EDSS po dvoletnem obdobju opazovanja. Njihov model je dosegel točnost 80 % in AUROC 80 % (5).

V tej predhodni raziskavi smo učili modele ML in DL in vrednotili ter medsebojno primerjali njihove sposobnosti napovedovanja prihodnjega povečanja EDSS iz MR-slik 393 bolnikov z MS.

## METODE

### Zajem podatkov raziskave

Poteka retrospektivna klinična raziskava, v kateri zbiramo podatke kliničnih in MR-preiskav bolnikov z MS iz štirih zdravstvenih ustanov. Soglasje Komisije Republike Slovenije za medicinsko etiko smo pridobili dne 16. 2. 2022 pod številko soglasja 0120-570/2021/5.

Raziskava trenutno vključuje 393 bolnikov z multiplo sklerozo (63 % žensk), starih 18–75 let, s povprečno starostjo 39,6 ± 10,3 let. Upoštevani so podatki 869 MR-preiskav glave s T1-poudarjenimi slikami in slikami, ki so nastale s tehniko slikanja

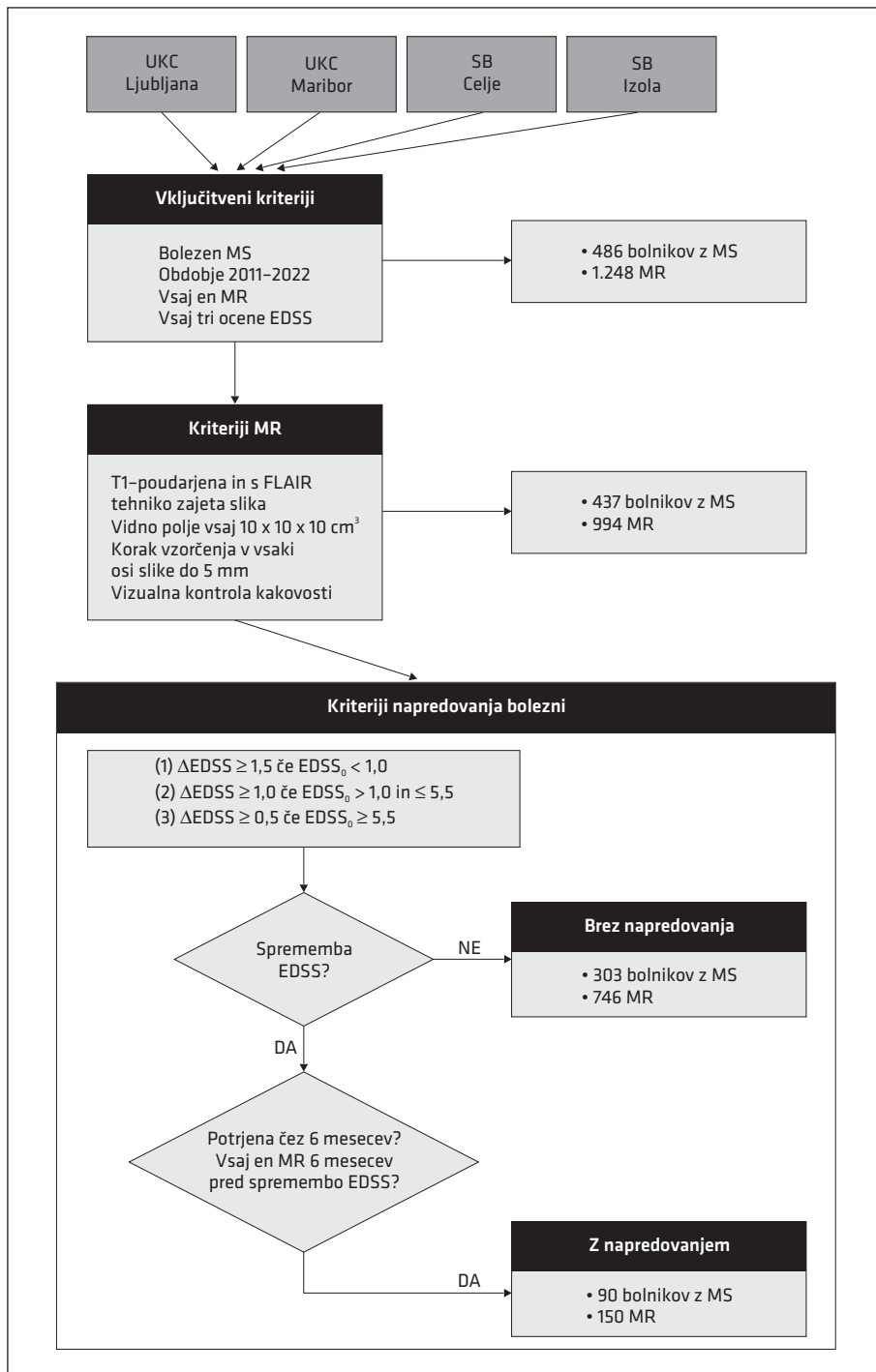
z zasičenjem signala iz tekočine (angl. *fluid attenuated inversion recovery*, FLAIR), zajetimi na petih različnih MR-napravah, in sicer Signa® HDxt (GE HealthCare™, Chicago, Illinois, ZDA), Achieva® (Philips Medical Systems® Nederland B.V., Best, Noord-Brabant, Nizozemska), Achieva dStream® (Philips Medical Systems® Nederland B.V., Best, Noord-Brabant, Nizozemska), MAGNETOM® Aera (Siemens®, München, Bavarska, Nemčija), MAGNETOM® Vida fit (Siemens®, München, Bavarska, Nemčija). Ob zajemu MR-slik je bil povprečni čas od diagnoze 7,5 ± 6,0 let. Vključitveni kriteriji za osebe v raziskavi, MR-preiskave in kriteriji napredovanja bolezni so podani v sliki 1.

### Obdelava magnetnoresonančnih slik

Vse T1-poudarjene in s FLAIR-tehniko zajete slike so bile sivinsko obnovljene in prostorsko poravnane glede na možganski atlas Montrealskega inštituta za nevrologijo (Montreal Neurological Institute, MNI 152) z afino preslikavo (6). Obdelane T1-poudarjene slike smo razmejili v 268 možganskih podstruktur z uporabo modela AssemblyNet in pridobili prostorninske deleže glede na ocenjeno prostornino lobanje (7). V slikah, zajetih s tehniko FLAIR, smo opravili razmejitev poškodb v beli možganovini z modelom DeepMedic in pridobili podatek o prostornini in številu teh poškodb (8).

### Napovedni modeli

Napovedne modele smo učili z namenom razvrščanja bolnikov z MS v dve skupini: v prvo s potrjenim prihodnjim povečanjem EDSS; torej z napredovanjem bolezni (angl. *with disease progression*, wDP) in v drugo za bolnike brez potrjenega povečanja EDSS; torej brez napredovanja bolezni (angl. *no disease progression*, nDP). Prostornine možganskih podstruktur, prostornine ter število poškodb (razbrano z MR-slik), demografske



**Slika 1.** Zajem podatkov raziskave. UKC – Univerzitetni klinični center, SB – splošna bolnišnica, MS – multipla skleroza, EDSS – razširjena lestvica stopnje prizadetosti (angl. *expanded disability status score*), FLAIR – zasičenje signala iz tekočine (angl. *fluid attenuated inversion recovery*).

podatke (starost, spol) in klinične podatke (izhodiščni EDSS, trajanje bolezni) smo uporabili za učenje modelov RF, SVM in Gaussove procese (GP). Učili smo tudi model DL vizualne geometrijske skupine (angl. *visual geometry group*, VGG), ki na vходу sprejme obdelani T1-poudarjeno sliko in s FLAIR-tehniko zajeto sliko (9). Vsi modeli vrnejo dvojiško odločitev wDP ali nDP.

Vsi modeli so bili učeni in vrednoteni neodvisno s štirikratnim prečnim preverjanjem, kjer so bili podatki v pregibih medsebojno primerljivi glede na delež MR-preiskav iz vsake ustanove, starost, razmerje med spoloma in razmerje bolnikov wDP ali nDP.

## REZULTATI

Kakovost napovedi z modeli smo ovrednotili z AUROC, točnostjo, občutljivostjo in specifičnostjo. Med VGG in ostalimi modeli smo s statističnimi testi ovrednotili razlike za posamezno mero, in sicer z DeLongovim testom za AUROC ter z McNemarjevim testom za točnost, občutljivost in specifičnost. Rezultati analize so navedeni v tabeli 1, o statistično značilni razliki pa poročamo pri mejni vrednosti  $\alpha = 0,05$  napake prve vrste.

Model DL VGG je presešel ostale tri modele ML glede na AUROC. Kot je razvidno iz tabele 1, sta modela GP in SVM dosegla višjo specifičnost, znatno nižji

vrednosti AUROC in občutljivosti, model SVM pa primerljivo točnost. Glede na nizko občutljivost in visoko specifičnost lahko opazimo, da sta modela GP in SVM razvrstila le majhen delež testnih primerov kot primere wDP. Model RF je dosegel najvišjo občutljivost (54 %), a hkrati najnižjo specifičnost (70 %). Skladno z najvišjim AUROC je model VGG dosegel najboljše razmerje med občutljivostjo (51 %) in specifičnostjo (83 %).

## RAZPRAVA

Naši rezultati AUROC, prikazani v tabeli 1, so v skladu z raziskavo Talonija in sodelavcev, kjer sta bili AUROC 72 % in 69 % doseženi za 2D-modele, naučene na koronalnih in aksialnih skladih petih rezin (4). Vendar naš model ni dosegel AUROC 81 % kot modeli v raziskavi Talonija in sodelavcev, naučeni na sagitalnih skladih 2D MR-rezin. Podobno velik vzorec bolnikov in arhitekturo modela VGG kot v pričujoči raziskavi so v svoji raziskavi uporabili Storelli in sodelavci, kjer so poročali o AUROC 80 %, točnosti 80 %, specifičnosti 90 % in občutljivosti 57 % (5). Naši rezultati, prikazani v tabeli 1, so nekoliko slabši v vseh merah vrednotenja, neskladje pa je lahko posledica razlik v skupini bolnikov, raznovrstnosti MR-slik (v naši raziskavi so bile slike posnete na štirih lokacijah) in predvsem v odstotku primerov

**Tabela 1.** Rezultati vrednotenja napovednih modelov napredovanja bolezni multiple skleroze. Navedene p-vrednosti so bile izračunane glede na rezultate modela vizualne geometrijske skupine. AUROC – površina površina pod krivuljo delovanja sprejemnika (angl. *area under receiver operating characteristic curve*), GP – Gaussovi procesi, RF – naključni gozdovi (angl. *random forests*), SVM – metoda podpornih vektorjev (angl. *support vector machines*), VGG – vizualna geometrijska skupina (angl. *visual geometry group*).

Model	AUROC	Točnost	Občutljivost	Specifičnost
GP	67 % <sup>a</sup>	76 % <sup>a</sup>	19 % <sup>a</sup>	92 % <sup>a</sup>
RF	62 % <sup>a</sup>	67 % <sup>a</sup>	54 % <sup>a</sup>	70 % <sup>a</sup>
SVM	63 % <sup>a</sup>	75 % <sup>b</sup>	11 % <sup>a</sup>	93 % <sup>a</sup>
VGG	74 %	76 %	51 %	83 %

<sup>a</sup> p < 0,001

<sup>b</sup> p = 0,325

wDP (33 % v raziskavi Storellija in sodelavcev in 23 % v naši raziskavi). Razlika je bila tudi v opredelitvi wDP in nDP; v raziskavi Storellija in sodelavcev so uporabili le dve MR-preiskavi znotraj dvoletnega intervala pred opredelitvijo wDP, naše opazovalno obdobje pa je bilo mnogo daljše (od 6 mesecev do 11 let pred opredelitvijo wDP) (5).

Ne glede na razlike je očitna nizka občutljivost tako našega modela VGG (51 %) kot modela v raziskavi Storelli in sodelavcev (57 %), kar pomeni, da modela uspešno prepoznata približno polovico bolnikov z MS, pri katerih kasneje pride do napredovanja bolezni (5).

O najvišji občutljivosti in AUROC so poročali Pinto in sodelavci z vrednostmi 76 in 86 %, vendar je primerjava z našimi rezultati omejena, saj smo opredelili napredovanje bolezni na podlagi ocene EDSS, medtem ko so Pinto in sodelavci uporabili klinično oceno kot merilo napredovanja in oceno EDSS za vhod modela. Poleg tega je njihov nabor podatkov obsegal le 21 bolnikov s potrjenim napredovanjem v SPMS, kar morda ni zadosten vzorec (3).

Za učinkovito zdravljenje bolnikov z MS, pri katerih obstaja večje tveganje za prihodnje napredovanje bolezni po EDSS ali v obliko SPMS, je ocena tega tveganja možna na podlagi analize MR-preiskav glave. Naši rezultati, ki temeljijo na modelu VGG, kažejo obetavno raziskovalno smer, pri čemer

je omenjeni napovedni model dosegel AUROC 74 %. Pri tem je pravilno prepoznal 51 % bolnikov wDP, medtem ko je bilo nDP pravilno prepoznanih 83 % bolnikov z MS. V celoti je bilo razvrščanje prihodnjega napredovanja bolezni pravilno pri 76 % bolnikov.

Rezultati naše raziskave kažejo, da uporabljene pristopi še niso primerni za klinično uporabo, za namene obvladovanja bolezni in/ali zdravljenja. Kljub temu menimo, da je razvoj načinov razvrščanja na podlagi slike, ki temelji na modelih DL iz MR-slik, obetavna raziskovalna smer. Prepoznavanje bolnikov z MS z velikim tveganjem za prihodnje napredovanje po EDSS oz. invalidnosti neposredno iz MR-preiskave je namreč enostavno za izvedbo, hkrati pa bi omogočilo pravočasno terapevtsko posredovanje, ki bi lahko preprečilo ali odložilo začetek napredovanja invalidnosti in prehoda v obliko SPMS.

## ZAHVALA

Raziskavo podpira podjetje Novartis Pharma Services Inc., podružnica Slovenija. Avtorji so imeli popoln nadzor nad vsebino in so sprejeli končno odločitev o vseh vidikih pričujočega prispevka.

Raziskavo podpira Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS), in sicer pod raziskovalnim programom št. P2-0232 in raziskovalnima projektoma št. J2-2500 in J2-3059.

## LITERATURA

1. Klineova S, Lublin FD. Clinical course of multiple sclerosis. *Cold Spring Harb Perspect Med.* 2018; 8 (9): a028928.
2. Branco D, Di Martino B, Esposito A, et al. Machine learning techniques for prediction of multiple sclerosis progression. *Soft Comput.* 2022; 26 (22): 12041–55.
3. Pinto MF, Oliveira H, Batista S, et al. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci Rep.* 2020; 10 (1): 21038.
4. Taloni A, Farrelly FA, Pontillo G, et al. Evaluation of disability progression in multiple sclerosis via magnetic-resonance-based deep learning techniques. *Int J Mol Sci.* 2022; 23 (18): 10651.
5. Storelli L, Azzimonti M, Gueye M, et al. A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. *Invest Radiol.* 2022; 57 (7): 423–32.
6. Fonov VS, Evans AC, McKinstry RC, et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage.* 2009; 47 (Suppl 1): S102.
7. Coupé P, Mansencal B, Clément M, et al. AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation. *Neuroimage.* 2020; 219: 117026.
8. Ravnik D, Jerman T, Pernuš F, et al. Dataset variability leverages white-matter lesion segmentation performance with convolutional neural network. In: *Medical Imaging 2018: Image Processing* [Internet]. 2018 [citirano 2023 Feb 13]. p. 388–96. Dosegljivo na: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10574/105741j/Dataset-variability-leverages-white-matter-lesion-segmentation-performance-with-convolutional/10.1117/12.2293702.full>
9. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv* [internet]. 2015 [citirano 2023 Feb 13]. Dosegljivo na: <http://arxiv.org/abs/1409.1556>