

Določitev optimalnih značilnic za zaznavo sprememb bele možganovine

Žiga Lesjak, Boštjan Likar, Franjo Pernuš, Žiga Špiclin

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška 25, 1000 Ljubljana, Slovenija
E-pošta: ziga.lesjak@fe.uni-lj.si

Povzetek. Postopki strojnega učenja omogočajo medsebojno primerjavo vpliva in določitev optimalnih značilnic za zaznavo sprememb lezij bele možganovine (LBM) ter s tem identifikacijo slikovne informacije, ki ključno vpliva na zmožnost zaznave teh sprememb. Z uporabo razvrščevalnika z naključnimi gozdovi smo na zbirki kliničnih magnetnoresonančnih (MR) slik 20 bolnikov, obolenih za multiplo sklerozo, ovrednotili in medsebojno primerjali pomembnost značilnic za zaznavo sprememb LBM s primerjavo z natančnimi referenčnimi obrisi. Določili smo nabor optimalnih značilnic, zanesljivost ocene pomembnosti značilnic in občutljivost na izbrani postopek razvrščanja. Ker izbor značilnic določa za zaznavo sprememb pomembne MR-sekvence in njihove nastavitve, ga lahko uporabimo za optimiziranje postopka zajema MR-slik. S tem lahko skrajšamo čas MR-preiskave in tako znižamo stroške ter zmanjšamo obremenitev bolnikov in medicinskega osebja. S predlagano metodologijo izbire optimalnih značilnic lahko načrtamo postopke zaznave sprememb z večjo občutljivostjo in specifičnostjo, kar v klinični praksi omogoča natančnejše spremljanje in boljše razumevanje poteka bolezni.

Ključne besede: izbor značilnic, zaznava sprememb, lezije bele možganovine, multipla sklerozo, obdelava slik

Selection of optimal features for white brain matter change detection

Machine learning algorithms can be used to determine the set of optimal features required for white matter lesion (WML) change detection. Identification of such features facilitates better understanding about which image information is the most important for the change detection. Using a random forest classifier, we evaluated and compared the importance of various image-derived features for the lesion change detection, assessed the reliability of the estimated feature importance and, finally, determined the ability to generalize the feature importance estimation to other classifiers. Evaluation was performed on a clinical database of magnetic resonance (MR) images of 20 Multiple Sclerosis patients by comparing them to corresponding reference WML changes delineations. The selection of optimal features may affect the setup of the MR acquisition parameters and the selection of sequences needed for the lesion change detection and as such can be used to optimize MR acquisition, for instance, to remove MR sequences not contributing to the change detection. Such optimizations may potentially decrease the cost and time required to acquire an MR patient study, a clear benefit for both the patients and medical personnel. The proposed feature selection methodology provides a basis to develop change detection methods with a high accuracy and specificity, which, in turn, would enable the clinicians a better insight into disease progression, increase their understanding of the underlying pathological processes and allow them to make a timely and fully-informed decision in case of ineffective treatments.
Keywords: feature selection, change detection, white matter lesions, multiple sclerosis, image analysis

1 UVOD

Proces avtomatske zaznave sprememb med dvema ali več magnetnoresonančnimi (MR) slikami glave vključuje primerjanje lastnosti solesnih struktur v teh slikah. Te lastnosti opišemo z značilnicami struktur, pri čemer je najosnovnejša značilnica sivinska vrednost. Značilnice so lahko tudi bolj kompleksne, kot na primer oddaljenost opazovanega slikovnega elementa od ventriklov, histogram sivinskih vrednosti v okolici opazovanega slikovnega elementa, na podlagi prostorsko poravnane atlasta določena *a priori* verjetnost, da opazovani slikovni element pripada določeni možganski strukturi, itn. Izbira primerne nabora značilnic je vselej pogojena s konkretnim problemom, ki ga želimo rešiti.

Pri načrtovanju postopkov za analizo in razgradnjo medicinskih slik se pogosto srečamo s tehničnimi vprašanji, povezanimi s predobdelavo in pripravo značilnic, kot na primer: Kakšen tip glajenja slik naj uporabim? Katere MR-sekvence naj uporabim? Ali naj slike predhodno preslikam v skupni prostor? V kakšnem vrstnem redu naj izvedem te postopke? Tudi s stališča klinične uporabe tehnologij analize slik se porajajo pomembna vprašanja: Kateri protokoli zajema slik so najustreznejši za poznejšo avtomatsko obdelavo? Kakšna naj bo prostorska ločljivost zajema slik? Pri majhnem začetnem naboru značilnic lahko optimalen nabor značilnic poiščemo na preprost način, in sicer s poizkušanjem vseh mogočih kombinacij značilnic. Z večanjem začetnega nabora značilnic pa tak pristop kaj hitro postane nepraktičen in računsko prezahteven,

saj se prostor iskanja optimalnih značilnic z večanjem začetnega nabora povečuje eksponentno. Iz teh razlogov se pri večjem začetnem naboru značilnic poslužujemo naprednejših postopkov za iskanje optimalnih značilnic.

Najpogosteje uporabljeni postopki za zaznavo sprememb v beli možganovini so tisti, ki temeljijo na analizi sivinskih vrednosti soležnih slikovnih elementov. Osnovi princip zaznave sprememb s temi postopki je vrednotenje podobnosti vrednosti ene ali več značilnic soležnih slikovnih elementov. V okviru tega dela bomo na primerih lezij bele možganovine (LBM) pri bolnikih, obolenih za multiplo sklerozo (MS), raziskali različne načine za ocenjevanje pomembnosti značilnic za zaznavanje sprememb bele možganovine. Kot osnovno orodje za določitev pomembnosti značilnic in izbor optimalnih značilnic bomo uporabili razvrščevalnik z naključnimi gozdovi (ang. *random forests*). Nato bomo z uporabo drugih razvrščevalnikov preverili zanesljivost z naključnimi gozdovi ugotovljene pomembnosti značilnic. Tako bomo ugotovili, ali dobljene zanesljivosti lahko pripišemo posameznim značilnicam in ali izbrani optimalni nabor značilnic na splošno lahko učinkovito uporabimo z drugimi razvrščevalniki.

1.1 Izbiranje značilnic

Izbiranje značilnic (ang. *feature selection*) je proces izbire optimalne podmnožice značilnic za čim boljše razvrščanje. Iz (začetnega) nabora z izbiranjem značilnic zmanjšamo dimenzijo prostora značilnic, ne da bi spremenili pomen značilnic. Ti postopki se razlikujejo od postopkov za znižanje dimenzionalnosti prostora značilnic, ki temeljijo na projekcijah (npr. metoda osnovnih komponent - ang. *principal component analysis*) ali kompresiji (npr. z uporabo teorije informacij). Izbiranje značilnic prinese številne prednosti, od katerih so najpomembnejše: 1. izognjenje prenasičenju (ang. *overfitting*) pri učenju razvrščevalnika in s tem povezano izboljšanje razvrščanja, 2. pohitritev in poenostavitev razvrščevalnika in 3. poglobljeno razumevanje pomembnosti značilnic, uporabljenih za graditev razvrščevalnika.

Iskanje optimalnih značilnic vnese v proces gradnje razvrščevalnika dodatno raven zahtevnosti. V nasprotju z gradnjo razvrščevalnika brez izbora značilnic, kjer le naučimo oziroma poiščemo optimalne parametre razvrščevalnika, moramo pri gradnji razvrščevalnika z izborom značilnic razširiti dimenzijo prostora iskanja z dodatnim podprostorom optimalnih značilnic. Tako moramo pri učenju razvrščevalnika poiskati optimalne parametre razvrščevalnika pri trenutni optimalni podmnožici značilnic, pri tem pa se moramo zavedati, da na splošno optimalni parametri razvrščevalnika s polnim naborom značilnic niso nujno enaki optimalnemu naboru značilnic pri razvrščevalniku, ki je zgrajen z optimalno podmnožico značilnic [2]. Postopke izbora značilnic delimo glede na to, kako poteka proces iskanja optimalnega podprostora značilnic [11].

Glede na način delovanja lahko postopke izbiranja značilnic razdelimo na: 1. postopke s filtriranjem, 2. z ovojem (ang. *wrapper methods*) in 3. z vgrajenim izborom (ang. *embedded methods*). Postopki s filtriranjem ocenijo ustreznost posameznih značilnic z intrinzičnimi lastnostmi podatkov – najpogosteje z razvrščanjem značilnic glede na neko oceno pomembnosti, pri čemer so najmanj pomembne nato odstranjene iz začetnega nabora značilnic. Prednosti teh postopkov so preprosta implementacija, računska nezahtevnost in dobro skaliranje za visokodimenzionalen začetni nabor značilnic. Taki postopki so neodvisni od razvrščevalnika, kar omogoča predhodno izbiranje optimalnih značilnic, s katerimi nato lahko preizkušamo različne razvrščevalnike. Tak pristop je hkrati tudi njihova največja pomanjkljivost, saj je izbor optimalnih značilnic na splošno vezan na uporabljeni razvrščevalnik. Dodatna slabost teh postopkov je tudi to, da so v veliki večini univariatni, kar pomeni, da pri izbiri optimalnih značilnic upoštevajo le eno značilnico hkrati in zanemarijo njihove medsebojne odvisnosti. Pozneje so bili razviti številni multivariatni postopki, ki na račun višje računske zahtevnosti upoštevajo tudi odvisnost med različnimi značilnicami.

Postopki za izbiro značilnic z ovojem delujejo tako, da problem iskanja parametrov razvrščevalnika “ovijejo” z dodatnim nivojem – iskanjem po prostoru vseh različnih mogočih podmnožic polnega nabora značilnic. Tako moramo za vsako mogočo podmnožico značilnic ponovno naučiti razvrščevalnik. Ker pri tem računska zahtevnost raste eksponentno s številom značilnic v začetnem naboru, se za iskanje optimalnega nabora značilnic v prostoru vseh podmnožic značilnic navadno uporablja hevristično iskanje. Glavni prednosti teh postopkov sta upoštevanje interakcije med podmnožicami značilnic in razvrščevalnikom ter upoštevanje medsebojnih odvisnosti značilnic. V primerjavi s postopki s filtriranjem so ti postopki izpostavljeni prenasičenju in računsko zelo zahtevni.

Tretji razred postopkov za izbiro značilnic so postopki z vgrajenim izborom, pri katerih je iskanje optimalnega nabora značilnic že vgrajeno v sam razvrščevalnik. Tako pri iskanju optimalnih parametrov razvrščevalnika hkrati iščemo tudi optimalno podmnožico značilnic. Prednost teh postopkov je v tem, da upoštevajo interakcijo med razvrščevalnikom in podmnožico značilnic, hkrati pa so računsko učinkovitejši kot postopki z ovojem. Pregled postopkov za izbiro značilnic skupaj z njihovimi prednostmi in slabostmi je prikazan v tabeli 1.

1.2 Izbor značilnic z uporabo naključnih gozdov

Naključni gozdovi so v zadnjih letih postali eno glavnih orodij tako pri strojnem učenju na splošno, kot tudi za specifične naloge, kot je obdelava medicinskih slik [3], [5], [7]. V literaturi so se naključni gozdovi za namen razvrščanja značilnic odrezali zelo dobro že na številnih problemih, tudi pri razvrščanju nelinearnih podatkov s kompleksnimi medsebojnimi interakcijami.

Tabela 1: Razvrstitev in pregled postopkov za izbiro značilnic ter njihove splošne prednosti in slabosti

Razred	Prednosti	Slabosti	Ime postopka
Filtriranje (univariatno)	Hitri, dobro skaliranje, neodvisni od razvrščevalnika	Ne upoštevajo medsebojnih odvisnosti značilnic	χ^2 , evklidska razdalja
Filtriranje (multivariatno)	Upoštevajo medsebojno odvisnost značilnic, neodvisni od razvrščevalnika	Računsko zahtevni, slabo skaliranje	Korelacijska izbira značilnic, <i>Markov blanket filter</i>
Postopki z ovajem	Preprosti, upoštevajo razvrščevalnik, upoštevajo medsebojne odvisnosti značilnic	Računsko zahtevni, občutljivi na prenasičenje	<i>Sequential forward selection, Beam search, Simulated annealing, Genetic algorithms</i>
Postopki z vgrajenim izborom	Upoštevajo razvrščevalnik, upoštevajo medsebojne odvisnosti značilnic	Izbor odvisen od izbranega razvrščevalnika	Odločitvena drevesa, uteženi naiven Bayesov razvrščevalnik, uteženi vektorski SVM

Naključni gozdovi spadajo med skupinske metode (ang. *ensemble methods*), ki za vsak problem razvrščanja uporabijo več posameznih odločitvenih dreves. Posamezna odločitvena drevesa so zgrajena na podlagi naključne podmnožice učenih podatkov (ang. *bootstrapping*). Tako je vsako drevo zgrajeno z različnim, naključno izbranim naborom značilnic. Razvrščanje oziroma napovedno vrednost, ki pomeni verjetnost razvrstitve z naključnimi gozdovi, dobimo tako, da združimo razvrstitve posameznih dreves v gozdu. V primerjavi z uporabo posameznih odločitvenih dreves se naključni gozdovi izkažejo za natančnejše in bolj zanesljive. Žal pa se njihove strategije razvrščanja v nasprotju s posamičnimi odločitvenimi drevesi ne da tako preprosto interpretirati. Pri posameznih odločitvenih drevesih lahko pomembnost posamezne značilnice hitro razberemo že iz njenega hierarhičnega položaja v odločitvenem drevesu, pri naključnih gozdovih pa moramo za določitev pomembnosti značilnic uporabiti druge pristope.

Najpreprostejši pristop za določitev pomembnosti značilnic v naključnem gozdu je preštevanje uporabe posamezne značilnice uporabljene prek vseh dreves v gozdu. Ta pristop na splošno ni najbolj primeren, saj moramo za zanesljivo oceno pomembnosti poleg pojavnosti posameznih značilnic v gozdu odločitvenih dreves upoštevati še: 1. število vzorcev, ki jih razvrsti posamezna razvejitev v drevesu, in 2. kakovost razvrščanja posamezne razvejitve. Naprednejši pristop za določitev pomembnosti značilnic dobimo, če s številom vzorcev uteženo povprečimo sposobnost ločevanja posameznih značilnic za vse razvejitve vseh dreves v gozdu [6]. Pri tem lahko sposobnost ločevanja merimo z izboljšanjem koeficienta Gini oziroma z zmanjšanjem entropije, od koder izhaja tudi ime ocene pomembnosti: pomembnost Gini (ang. *Gini Importance - GI*).

Najnaprednejša mera za izračun pomembnosti značilnic s pomočjo naključnih gozdov je tako imenovana *pomembnost točnosti s permutacijo* (ang. *Permutation Accuracy Importance - PAI*). Za oceno pomembnosti po tem postopku potrebujemo dodatno validacijsko zbirko podatkov X. Nato

posamezen stolpec, ki pomeni eno značilnico v vhodnih podatkih X_j , naključno premešamo in tako izničimo njen vpliv na napovedno vrednost Y. Pomembnost posamezne permutirane značilnice nato določimo tako, da podatke s permutirano značilnico uporabimo na naključnem gozdu in opazujemo spremembo napovednih vrednosti Y. Pomembnost permutirane značilnice lahko tako definiramo kot relativno razliko v točnosti napovednih vrednosti Y pred permutacijo in po njej, pri čemer lahko za definicijo točnosti izberemo različne validacijske metrike. Ta postopek ponovimo za vsako posamezno značilnico, katere pomembnost želimo oceniti. Prednost postopka PAI je, da ni specifičen za naključne gozdove, in ga lahko zato uporabljamo za oceno pomembnosti značilnic različnih razvrščevalnikov.

2 MATERIALI IN METODE

Za določitev pomembnosti značilnic za zaznavo sprememb bele možganovine smo uporabili postopek z vgrajenim izborom značilnic, ki temelji na razvrščanju vzorcev z naključnimi gozdovi. Le-ta nam namreč že po svoji zasnovi omogoča določanje pomembnosti posameznih značilnic.

2.1 Validacijska zbirka slik

Za učenje in validacijo naučenega naključnega gozda smo uporabili javno objavljeno longitudinalno zbirko MR-slik z referenčnimi obrisi sprememb LBM [8]. Zbirka vsebuje referenčne in kontrolne MR-preiskave 20 MS bolnikov. Slike so bile zajete na 1,5T Philips MR-napravi v Univerzitetnem kliničnem centru v Ljubljani. Vsaka MR-preiskava vsebuje 2D T1w (*spin echo sequence*, TR=600 ms, TE=15 ms, FA=90°, vzorčenje 0,9×0,9×3 mm brez razmika med rezinami), 2D T2w (*spin echo sequence*, TR=4500 ms, TE=100 ms, FA=90°, vzorčenje 0,45×0,45×3 mm brez razmika med rezinami) in 2D FLAIR sliko (TR=11000, TE=140, TI=2800, FA=90, vzorčenje 0,9×0,9×3 mm brez razmika med rezinami).

V tabeli 2 so prikazani demografski podatki o bolnikih, njihovo bolezensko stanje in trenutna terapija. Čas med zaporednimi MR-preiskavami bolnikov je bil od 81 do 723 dni (mediana: 311, IQR: 223). Vsi bolniki so podali pisno soglasje za uporabo slik za namen tega dela, njihove slike pa so bile pred računalniško obdelavo anonimizirane. Referenčni obrisi sprememb LBM so bili zgrajeni na podlagi slike razlik in so kot taki zaradi visoke občutljivosti dobro orodje za validacijo postopkov zaznave sprememb LBM [8].

Slike iz zbirke smo psevdonaključno ločili v dve skupini, pri čemer je testna skupina vsebovala 66 % vseh slik, validacijska skupina pa preostalih 33 % slik. Pri razporejanju slik bolnikov smo pazili, da sta bili obe skupini uravnoteženi v smislu enake porazdelitve skupnega volumna LBM pri posameznih bolnikih.

2.2 Začetni nabor značilnic

Iz vsakega para dveh zaporednih MR-preiskav smo izločili naslednji nabor začetnih značilnic: a) šest enostavnih značilnic sivinskih vrednosti referenčnih in kontrolnih T1w, T2w in FLAIR slik, b) dve značilnici oddaljenosti opazovanega slikovnega elementa od stene ventriklov oziroma možganske skorje in c) sedem značilnic, ki so pomenile vse možne kombinacije slike podobnosti STI različnih MR-sekvenc [8]:

$$DM_{STI}(\mathbf{x}_i) = \|\mathcal{F}_M(\mathbf{x}_i) - \mathcal{B}_M(\mathbf{x}_i)\|, \quad (1)$$

pri čemer se \mathcal{B}_M in \mathcal{F}_M nanašata na predobdelano referenčno oziroma kontrolno MR-sliko sekvence \mathcal{M} , \mathbf{x}_i je sivinska vrednost slikovnega elementa i , $\|\cdot\|$ pa L2 norma razlik istoležnih slikovnih elementov. Skupaj smo iz vsakega para zaporednih MR-preiskav izločili 15 značilnic.

Začetni nabor preprostih in pomenljivih značilnic nam omogoča preprosto vrednotenje in interpretacijo rezultatov, zaradi njihovega relativno nizkega števila pa je obvladljiva tudi računska zahtevnost.

2.3 Učenje naključnega gozda

Pri učenju naključnega gozda moramo izbrati številne parametre, kot so: število odločitvenih dreves v gozdu, največja globina odločitvenih dreves, število vzorcev, upoštevanih pri vsakem razcepu odločitvenega drevesa, itn. V našem primeru smo začetne parametre izbrali na podlagi relevantne literature, in sicer take, ki so se pri podobnih problemih izkazali za ustrezne [1], [9]. Parametre smo v poskusih prilagodili. Ustreznost izbranih parametrov smo kvalitativno preverili z opazovanjem Diceovega podobnostnega koeficienta (DSC – ang. *Dice similarity coefficient*) razvrščevalnika na učni in validacijski zbirki slik pri različnih vrednostih parametrov (število dreves, maksimalna globina dreves itd.), med učenjem in po njem. V poskusih uporabljeni parametri so prikazani v tabeli 3.

Naključni gozd smo učili postopoma, in sicer tako, da smo za vsakega od n bolnikov naučili po k odločitvenih

dreves. Celoten gozd je tako obsegal $n \times k$ odločitvenih dreves. Ker je lahko vseh dreves veliko in s tem učenje računsko zelo zahtevno, hkrati pa lahko pride do prenasičenja, smo po končanem učenju obdržali le t po kakovosti razvrščanja najboljših dreves, preostala pa smo zavrgli. Za določitev najboljših odločitvenih dreves smo uporabili vzorce iz slik istih bolnikov kot za učenje, le da vzorcev nismo decimirali, za kriterij zmožnosti oziroma kakovosti razvrščanja dreves pa smo uporabili metriko DSC. Ustreznost modela smo na koncu ovrednotili z izračunom občutljivosti (ang. *true positive rate* - *TPR*), nespecifičnosti (ang. *false positive rate* - *FPR*) in metriko DSC na validacijski zbirki slik.

Pri učenju naključnih gozdov moramo biti pozorni na razpršenost vzorcev in razmerje med številom vzorcev v posameznih razredih. Ker je prostornina sprememb LBM v primerjavi s prostornino celotne bele možganovine izjemno majhna, smo morali posebno pozornost nameniti tudi ustrezni izbiri učnih vzorcev pri posameznem bolniku. Izkaže se, da naključni gozdovi delujejo najbolje, če sta obe skupini vzorcev enakomerno zastopani. Zato smo pri učenju razvrščevalnika poleg celotnega nabora pozitivnih vzorcev (tj. LBM) za vsakega bolnika izbrali še enako število negativnih vzorcev (z območja zdrave bele možganovine). Da bi zajeli večji del variabilnosti negativnih vzorcev, smo za vsakega bolnika vzorce vzorčili večkratno, pri čemer smo pozitivne vzorce preprosto pomnogoterili, medtem ko so bili negativni vzorci vselej izbrani naključno. Ker so bili vzorci normalizirani že kot del postopka predobdelave pri pripravi zbirke slik [8], končnih izbranih vzorcev za učenje izbora značilnic in učenje razvrščevalnika nismo dodatno normalizirali.

2.4 Analiza zanesljivosti in posplošljivosti

S pomočjo naučenega in validiranega naključnega gozda smo nato opazovali pomembnost posameznih značilnic z dvema različnima postopkoma: GI in PAI. Zanimalo nas je, kako stabilne oziroma zanesljive so tako pridobljene ocene pomembnosti značilnic, zato smo naključni gozd naučili večkrat, pri čemer smo pri posameznem procesu učenja odstranili po eno značilnico iz celotnega začetnega nabora in s pomočjo korelacije opazovali, kako se kot posledica odstranitve značilnice spremeni ocena pomembnosti preostalih značilnic.

Ker se v praksi pogosto uporabljajo različni razvrščevalniki, smo v nadaljevanju preverili, ali se ugotovljene pomembnosti značilnic z uporabo naključnih gozdov lahko posplošijo na druge razvrščevalnike. Vprašali smo se: "Ali so značilnice, ki so pomembne za razvrščanje z uporabo naključnih gozdov, enako pomembne tudi, če uporabimo druge postopke razvrščanja?". Da bi odgovorili na to vprašanje, smo na enaki učni zbirki slik naučili nekatere od pogosto uporabljenih razvrščevalnikov: kvadratična diskriminativna analiza (ang. *Quadratic Discriminant Analysis* – *QDA*), Adaboost in naivni Bayes (ang. *Naive Bayes*).

Tabela 2: Demografski podatki bolnikov in trenutna terapija

Spol	Starost	Tip bolezni MS	Terapija
			2 brez zdravil
			6 Gilenya
14 žensk	od 19 do 50 let	1 sekundarno–progresivna	2 Copaxon
6 moških	mediana: 34 let	14 recidivno–remitentna	2 Tysabri
		5 neznano	1 Tecfidera
			1 Aubagio
			6 neznano

Tabela 3: Parametri, uporabljeni pri graditvi naključnih gozdov

Opis parametra	Vrednost	Omejitve
št. odločitvenih dreves na bolnika (k)	št. značilnic	min 15, maks 40
št. značilnic, upoštevanih za razcep	$\sqrt{\text{št. značilnic}}$	
maksimalna globina odločitvenih dreves	$\sqrt{\text{št. značilnic}}$	min 3
končno število odločitvenih dreves	150	
min. št. vzorcev za razcep	300	
min. št. vzorcev v listu	50	

Tabela 4: Metrike kakovosti razgradnje slik z naključnim gozdom na učni in validacijski zbirki slik.

Zbirka slik	TPR [†]	FPR [‡]	DSC
Učna	0,834	0,055	0,882
Validacijska	0,823	0,080	0,863

[†] občutljivost, [‡] specifičnost

Zanimalo nas je tudi, kako sprememba parametrov razvrščevalnika vpliva na oceno pomembnosti značilnic, zato smo ponovno naučili nov naključni gozd, tokrat s spremenjenimi parametri glede na tabelo 3: maksimalna globina drevesa 5, minimalno število vzorcev za razcep 2, minimalno število vzorcev v listu 1 in 150 dreves v gozdu, pri čemer so bila drevesa izračunana na podlagi slik vseh bolnikov. Za vse naučene razvrščevalnike smo izračunali metriko pomembnosti značilnic PAI, pri čemer smo za kakovost razvrščanja uporabili metriko DSC. S korelacijo PAI novih razvrščevalnikov in originalnega naključnega gozda pa smo primerjali, ali so ocene pomembnosti, pridobljene z naključnim gozdom, prenosljive na druge razvrščevalnike.

3 REZULTATI

Tabela 4 kaže metrike, ki jih naučeni naključni gozd dosegajo na validacijski zbirki slik. Vidimo, da se rezultati med učno in validacijsko zbirko slik le malo razlikujejo, kar kaže na to, da je naključni gozd dobro naučen in da ne prihaja do občutnega prenasičenja.

V tabeli 5 in na sliki 1 so prikazane vrednosti metrik GI in PAI (mediana in IQR), ki vrednotita pomembnost

posameznih značilnic. Mediana in IQR sta bili uporabljeni namesto srednje vrednosti in standardne deviacije, ker je bilo mogoče že kvalitativno oceniti, da porazdelitve vrednosti GI in PAI niso normalne. Izračun mediane in IQR je bil narejen na posameznih razredih vzorcev iz validacijske zbirke.

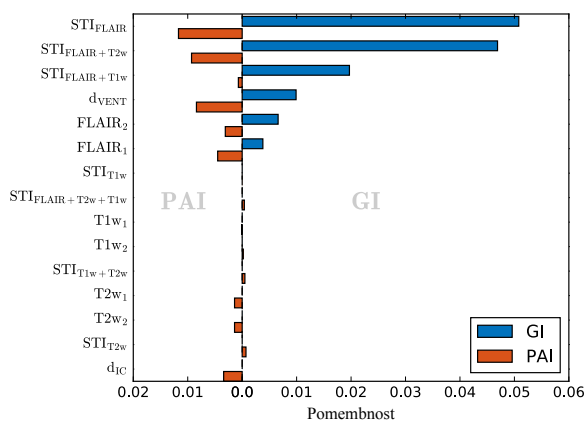
Oba postopka za določitev pomembnosti kot najboljšo značilnico izbereta STI_{FLAIR} , kar se ujema z ugotovitvami v strokovni literaturi, kjer se STI_{FLAIR} pogosto uporablja za zaznavanje sprememb LBM [4], [10]. Napovedi GI in PAI se, kljub določenim odstopanjem dobro ujemata tudi za druge značilnice (korelacija Pearson $r = 0,82$). Opazimo tudi, da pri določenih značilnicah PAI dosega negativne vrednosti. Pri teh značilnicah je kljub skrbni izbiri parametrov in validaciji prišlo do rahlega prenasičenja razvrščevalnika. Posledično se rezultati razgradnje sprememb LBM in s tem vrednost validacijske metrike izboljšajo, ko pri izračunu PAI “izklopimo” to značilnico.

Stabilnost ocene pomembnosti značilnic GI in PAI, določena kot korelacija GI/PAI med naključnim gozdom, ki je bil naučen na polni učni zbirki, in naključnim gozdom, ki je bil naučen na učni zbirki, ki smo ji odstranili določeno značilnico, je prikazana v tabeli 6. Iz rezultatov vidimo, da je GI izjemno stabilna, saj znaša njena povprečna korelacija 0,991 in se ne glede na to, katero značilnico odstranimo iz nabora začetnih značilnic, le malo spremeni (minimalna vrednost korelacije $r_{min} = 0,978$). Pomembnost z metriko PAI izkazuje nižjo stabilnost (povprečna korelacija $\bar{r} = 0,822$) in večjo variabilnost le-te, pri čemer je padec korelacije s PAI večji za značilnice, ki so pomembnejše. Med padcem korelacije s PAI in pomembnostjo značilnice

Tabela 5: Ocene pomembnosti značilnic glede na metriki GI in PAI. Značilnice so urejene glede na padajoče vrednosti mediane GI.

Ime značilnice	Mediana GI ($\times 10^{-2}$)	GI IQR ($\times 10^{-2}$)	Mediana PAI ($\times 10^{-2}$)	PAI IQR ($\times 10^{-2}$)
STI _{FLAIR}	5,08	29,01	1,17	3,35
STI _{FLAIR+T2w}	4,69	29,23	0,93	0,59
STI _{FLAIR+T1w}	1,97	14,64	0,07	1,23
d _{VENT}	0,99	5,80	0,84	0,85
FLAIR ₂	0,66	4,25	0,31	0,54
FLAIR ₁	0,38	3,27	0,45	0,53
STI _{T1w}	0,00	0,00	0,00	0,02
STI _{FLAIR+T2w+T1w}	0,00	0,04	-0,04	0,13
T1w ₁	0,00	0,13	0,01	0,08
T1w ₂	0,00	0,26	-0,02	0,10
STI _{T1w+T2w}	0,00	0,63	-0,05	0,16
T2w ₁	0,00	0,80	0,14	0,23
T2w ₂	0,00	1,05	0,14	0,27
STI _{T2w}	0,00	1,36	-0,07	0,35
d _{IC}	0,00	1,46	0,34	0,19

d_{VENT} – oddaljenost od ventriklov, d_{IC} – oddaljenost od roba možganov



Slika 1: Pomembnosti značilnic glede na metriki GI in PAI. Značilnice so urejene glede na padajoče vrednosti mediane GI.

po GI lahko namreč opazimo relativno visoko negativno korelacijo $r = -0,798$.

Predno lahko analiziramo zmožnost prenosa pridobljenih ocen pomembnosti značilnic na druge razvrščevalnike, je le-te treba ustrezno naučiti. Uspešnost učenja razvrščevalnikov smo preverili na enak način kot pri učenju originalnega naključnega gozda. Relativne razlike naučenih razvrščevalnikov med učnim in validacijskim setom niso presegale 4 % glede na DSC. Za vsak posamezni razvrščevalnik smo s pomočjo validacijske zbirke izračunali tudi PAI, ki smo ga s pomočjo korelacije primerjali s PAI in GI iz naključnega gozda. Iz tabele 7 in slike 2 je razvidno, da se različni razvrščevalniki kljub temu, da dosegajo primerljive zmožnosti v smislu validacijskih metrik (npr. DSC), zanašajo na različne kombinacije značilnic.

Tabela 6: Stabilnost ocene pomembnosti značilnic glede na metriki GI in PAI. Večja vrednost korelacije kaže na večjo stabilnost.

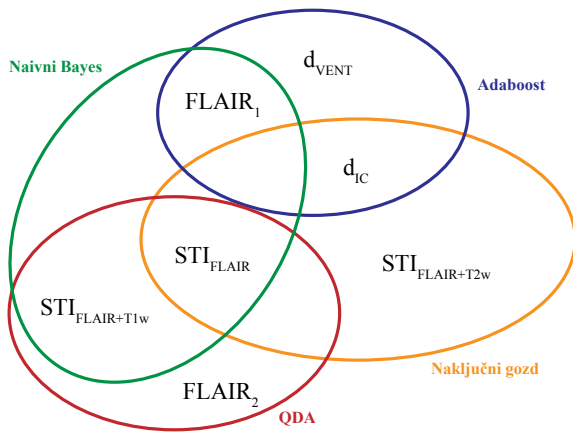
Odstranjena značilnica	Korelacija z GI	Korelacija s PAI
T1w ₁	0,983	0,819
T1w ₂	0,988	0,799
T2w ₁	0,988	0,826
T2w ₂	0,996	0,848
FLAIR ₁	0,995	0,867
FLAIR ₂	0,986	0,858
STI _{T1w}	0,997	0,833
STI _{T2w}	0,996	0,849
STI _{FLAIR}	0,978	0,589
STI _{T1w+T2w}	0,990	0,868
STI _{FLAIR+T1w}	0,997	0,862
STI _{FLAIR+T2w}	0,996	0,729
STI _{FLAIR+T2w+T1w}	0,994	0,819
d _{VENT}	0,995	0,890
d _{IC}	0,991	0,876
Povprečje	0,991	0,822

d_{VENT} – oddaljenost od ventriklov

d_{IC} – oddaljenost od roba možganov

Tabela 7: Korelacija PAI različnih razvrščevalnikov s PAI in GI naključnega gozda.

Razvrščevalnik	Korelacija s GI	Korelacija s PAI
QDA	0,332	0,287
Adaboost	-0,042	0,470
Naivni Bayes	0,550	0,543
Naključni gozd	0,096	0,535



Slika 2: Tri najpomembnejše značilnice glede na metriko PAI za vsakega od štirih različnih razvrščevalnikov: naivni Bayes, Adaboost, QDA in naključni gozd.

4 RAZPRAVA

V okviru tega poglavja smo proučili enega od računsko relativno nezahtevnih postopkov za določanje pomembnosti značilnic za zaznavo sprememb LBM iz longitudinalnih MR-preiskav glave. Pri tem smo se osredinili zlasti na stabilnost oziroma zanesljivost postopka izbora optimalnih značilnic ter prenosljivost teh značilnic na druge postopke razvrščanja. Splošen in zanesljiv postopek za določitev pomembnih značilnic za zaznavo sprememb LBM bi namreč pripomogel k boljšemu razumevanju problema oziroma kritične informacije za zaznavo teh sprememb, kar bi lahko posledično privedlo do novih, izboljšanih postopkov zaznave sprememb LBM ter do bolj optimalnega izbora in nastavitve slikovnih sekvenc zajema MR-preiskave.

Primerjali smo dve različni metriki (GI in PAI) za vrednotenje pomembnosti značilnic. S tem namenom smo naučili naključni gozd na naboru petnajstih značilnic, ki so bile v relevantni literaturi pogosto uporabljene v postopkih za zaznavo sprememb LBM, temelječih na primerjavi sivinske informacije: intenzitete referenčnih in kontrolnih T1w, T2w in FLAIR slikovnih sekvenc ter intenzitete različnih kombinacij STI (slika podobnosti sekvenc). Dodali pa smo tudi značilnice o anatomskem položaju na podlagi oddaljenosti opazovanega slikovnega elementa od ventriklov in od roba možganske skorje.

Analiza pomembnosti značilnic je kot najpomembnejšo značilnico za zaznavanje sprememb LBM pokazala STI_{FLAIR}, kot pomembni pa sta se izkazali tudi STI_{FLAIR+T2w} in razdalja med opazovanim slikovnim elementom in ventrikli. Te ugotovitve sovpadajo z dosedanjimi ugotovitvami nevro radiologov in drugih strokovnjakov, saj se razlike v lezijah najlepše opazijo na STI-sliki podobnosti sekvenc FLAIR ali T2w, najpogosteje pa se pojavljajo v neposredni okolici ventriklov [10].

Oceno pomembnosti z GI dobimo inherentno že iz naučenega naključnega gozda, medtem ko PAI izračunamo s pomočjo dodatne validacijske zbirke in je kot taka uporabna tudi za oceno pomembnosti značilnic drugih razvrščevalnikov. Pomembnost GI se izkaže kot izjemno robustna na rahle spremembe v naboru uporabljenih značilnic, medtem ko PAI izkazuje rahlo nižjo robustnost, zlasti če iz nabora značilnic odstranimo pomembnejše značilnice. Kljub temu je korelacija med ocenami pomembnosti GI in PAI, pridobljenimi z originalnim naključnim gozdom in naključnimi gozdovi, ki so jim bile odstranjene specifične značilnice, visoka in v povprečju dosega 0,991 za GI in 0,822 za PAI.

Ugotovili smo, da neposredna posplošitev izbora optimalnih značilnic z uporabo naključnih gozdov na druge razvrščevalnike ni mogoča. Primerjava z drugimi razvrščevalniki je pokazala, da se pomembnost posameznih značilnic spreminja od razvrščevalnika do razvrščevalnika (slika 2). Do razlik v oceni pomembnosti značilnic pa lahko pride že v posameznih razvrščevalnikih, kot smo to pokazali na primeru naključnih gozdov. S spreminjanjem parametrov razvrščevalnika namreč hitro dobimo drugačno oceno pomembnosti značilnic, zato moramo biti pri učenju razvrščevalnika previdni ter ga vselej dobro validirati.

Eden od ciljev določitve pomembnosti značilnic za longitudinalno zaznavo sprememb LBM je vsekakor tudi izboljšanje oziroma optimizacija zajema MR-slik v klinični praksi. V zadnjem času se namreč radiologi na klinikah vse pogosteje srečujejo s problemom izbire ustreznih MR-sekvenc in njihovih parametrov za čim bolj optimalen zajem slik bolnika. Katere in koliko različnih sekvenc je treba zajeti, kakšno naj bo vidno polje in kako tanke naj bodo rezine zajetih 3D MR-slik? Vsi ti parametri zajema so kompromis med časom zajema slik in njihovo kakovostjo. V klinični praksi še danes zajete MR-slike pogosto pregledujejo ročno, rezino za rezino. Za tak način pregledovanja se slike navadno zajemajo v debelejših rezinah (npr. 3 mm ali celo 5 mm), saj radiolog v praksi zaradi časovnih omejitev stežka pregleda slike z večjim številom rezin.

Kljub temu pa zadnje klinične smernice za zaznavanje sprememb LBM iz MR-slik priporočajo zajem 3D izotropnih MR-slik s prostorsko ločljivostjo 1 mm [12]. V literaturi namreč poročajo, da izotropno zajete MR-slike s prostorsko ločljivostjo 1 mm ali manj občutno pripomorejo k natančnosti avtomatske analize slik. Sklepamo lahko, da bi slike višje ločljivosti izboljšale tudi postopke zaznave sprememb LBM. Zanimivo bi bilo ovrednotiti vpliv ločljivosti slik na zaznavanje sprememb LBM, za kar bi potrebovali razširjeno zbirko slik, ki vsebuje slike, zajete pri različnih prostorskih ločljivostih. Če bi se izkazalo, da so za avtomatsko zaznavo sprememb LBM potrebne slike višje ločljivosti, ki za zajem potrebujejo dalj časa, bi bilo smiselno proučiti tudi, ali lahko katero od MR-sekvenc opustimo. Za rutinsko spremljanje bolnikov z MS se namreč uporabljajo številne MR-

sekvence, kot so T1- in T2-utežene slike, T1-utežene slike s kontrastom, FLAIR, DWI itn. Rezultati analize pomembnosti posameznih slikovnih sekvenc kažejo, da je za avtomatsko zaznavanje sprememb LBM daleč najpomembnejša sekvenca FLAIR, posredno pa tudi T1-utežena sekvenca, saj se uporablja za razgradnjo zdravih možganskih struktur (npr. možganska skorja, ventrikli). Zajem preostalih sekvenc bi lahko opustili in s tem skrajšali čas, ki ga potrebujemo za kontrolno slikanje MS bolnika, kar bi lahko zmanjšalo tudi s tem povezane stroške obravnave teh bolnikov.

Treba pa se je zavedati, da optimalni protokol zajema MR-slik za zaznavo sprememb LBM nikakor ni zadosten za rutinsko spremljanje tako kompleksne bolezni, kot je MS. LBM so namreč le eden izmed pomembnih slikovnih biomarkerjev bolezni MS. Za druge biomarkerje, kot so na primer lezije ob možganski skorji ali pa meritve atrofije posameznih predelov možganov, bi bil optimalen izbor MR-sekvenc za avtomatsko analizo gotovo drugačen. Za optimalen nabor protokolov za MR-preiskavo bolnikov z MS bi tako morali oceniti vpliv različnih MR-sekvenc na postopke za zaznavo sprememb številnih različnih pomembnih slikovnih biomarkerjev ter pri tem upoštevati časovne in finančne omejitve, s katerimi se srečujemo v klinični praksi.

ZAHVALA

Raziskavo je omogočila Javna agencija za raziskovalno dejavnost Republike Slovenije v okviru projektov J7-6781, J2-7211 in J2-7118.

LITERATURA

- [1] Antonio Criminisi and J. Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Science & Business Media, January 2013. Google-Books-ID: nUhDAAAQBAJ.
- [2] Walter Daelemans, Véronique Hoste, Fien De Meulder, and Bart Naudts. Combined Optimization of Feature Selection and Algorithm Parameters in Machine Learning of Language. In Nada Lavrač, Dragan Gamberger, Hendrik Blockeel, and Ljupčo Todorovski, editors, *Machine Learning: ECML 2003*, number 2837 in Lecture Notes in Computer Science, pages 84–95. Springer Berlin Heidelberg, September 2003. DOI: 10.1007/978-3-540-39857-8_10.
- [3] Ramón Díaz-Uriarte and Sara Alvarez de Andrés. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3, 2006.
- [4] Colm Elliott, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. *IEEE transactions on medical imaging*, 32(8):1490–1503, August 2013.
- [5] Colm Elliott, Simon J. Francis, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Bayesian classification of multiple sclerosis lesions in longitudinal MRI using subtraction images. In *Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention: Part II, MICCAI'10*, pages 290–297, Berlin, Heidelberg, 2010. Springer-Verlag.
- [6] Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.
- [7] A. Geert Heidema, Jolanda MA Boer, Nico Nagelkerke, Edwin CM Mariman, Daphne L. van der A, and Edith JM Feskens. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. *BMC Genetics*, 7:23, 2006.
- [8] Žiga Lesjak, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. Validation of White-Matter Lesion Change Detection Methods on a Novel Publicly Available MRI Image Database. *Neuroinformatics*, May 2016.
- [9] Andy Liaw and Matthew Wiener. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002.
- [10] Bastiaan Moraal, Dominik S. Meier, Peter A. Poppe, Jeroen J. G. Geurts, Hugo Vrenken, William M. A. Jonker, Dirk L. Knol, Ronald A. van Schijndel, Petra J. W. Pouwels, Christoph Pohl, Lars Bauer, Rupert Sandbrink, Charles R. G. Guttmann, and Frederik Barkhof. Subtraction MR Images in a Multiple Sclerosis Multicenter Clinical Trial Setting. *Radiology*, 250(2):506–514, February 2009.
- [11] Yvan Saeyns, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in. *Bioinformatics*, 23(19):2507–2517, October 2007.
- [12] H. Vrenken, M. Jenkinson, M. A. Horsfield, M. Battaglini, R. A. van Schijndel, E. Rostrup, J. J. G. Geurts, E. Fisher, A. Zijdenbos, J. Ashburner, D. H. Miller, M. Filippi, F. Fazekas, M. Rovaris, A. Rovira, F. Barkhof, N. de Stefano, and MAGNIMS Study Group. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. *Journal of Neurology*, 260(10):2458–2471, October 2013.

Žiga Lesjak se je leta 2012 diplomiral na Fakulteti za elektrotehniko Univerze v Ljubljani, kjer se je po diplomi zaposlil kot mladi raziskovalec. Njegovo področje raziskovanja obsega obdelavo in analizo biomedicinskih slik s poudarkom na postopkih avtomatske zaznave sprememb iz magnetnoresonančnih slik glave.

Boštjan Likar se je rodil leta 1971. Na Fakulteti za elektrotehniko v Ljubljani je diplomiral leta 1995, magistriral leta 1998 ter doktoriral leta 2000. Istega leta je doktoriral tudi na Medicinski fakulteti v Utrechtu na Nizozemskem. Leta 1996 je bil izvoljen v naziv asistent, leta 1999 v naziv višji predavatelj, leta 2001 v naziv docent, leta 2006 v naziv izredni profesor in leta 2011 v naziv redni profesor na Fakulteti za elektrotehniko Univerze v Ljubljani. Njegovo raziskovalno delo v Laboratoriju za slikovne tehnologije obsega vizualno kontrolo kakovosti, sisteme s strojnimi in računalniškimi vidom ter obdelavo in analizo biomedicinskih in hiperspektralnih slik.

Franjo Pernuš je na Fakulteti za elektrotehniko v Ljubljani diplomiral leta 1976, magistriral leta 1979 ter doktoriral leta 1991. Od leta 1976 je zaposlen na Fakulteti za elektrotehniko v Ljubljani, kjer je redni profesor in vodja Laboratorija za slikovne tehnologije. Njegovo raziskovalno področje obsega računalniški vid, obdelavo medicinskih slik ter uporabo metod obdelave slik in razpoznavanja vzorcev pri različnih biomedicinskih in industrijskih problemih.

Žiga Špiclin je leta 2011 doktoriral na Fakulteti za elektrotehniko Univerze v Ljubljani, kjer je trenutno zaposlen kot raziskovalec. Ukvarja se z razvojem in vrednotenjem numeričnih postopkov za prostorsko poravnavo slik, izboljšanje kakovosti slik in razgradnjo slik za uporabo v biomedicini in aplikacijah računalniškega vida.