



Fakultet tehničkih nauka
Novi sad 2023

Izdvajanje obeležja i klasifikacija sa MRI snimaka

Aleksandra Zarija E1 98/2022

Teodora Jezdimirović E1 99/2022

Baza podataka

MRI snimci dojki 01

922 pacijenta 03

iz jedne institucije 02

više serija za svakog
pacijenta 04

Pravljenje dodatnih foldera



Folder koji sadrži slike iz svake serije za svakog pacijenta na kojima su lezije vidljive



Folder koji sadrži maske napravljene na osnovu slika iz foldera 01 i koordinata lezija iz excel fajla
Annotation Boxes



Folder koji sadrži maske koje su vertikalno simetrične onima iz foldera 02



Lezije vs zdravo tkivo

Radiomics obeležja tražimo na mestima gde postoje lezije i gde je zdravo tkivo.



Lezije pronalazimo pomoću maski iz foldera 2, a zdravo tkivo pomoću maski iz foldera 3 (samo u slučajevima kada ne postoji bilateralni kancer).





Radiomics obeležja

Pomoću PyRadiomics paketa izdvajamo obeležja na osnovu kojih vršimo klasifikaciju lezija na kancerogene i nekancerogene.

Korišćena su sledeća radiomics obeležja: First order features, Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone (GLSZM), Gray Level Run Length Matrix (GLRLM).



Radiomics obeležja

Statistike prvog reda (First order features) opisuju distribuciju intenziteta vokselu unutar regiona slike definisanog maskom kroz uobicajene i osnovne metrike. ○

Tu spadaju: energija, entropija, srednja vrednost, minimum, maksimum, medijan, standardna devijacija, varijansa... ○

Radiomics obeležja

Gray Level Co-occurrence Matrix (GLCM) opisuje zajedničku funkciju verovatnoće drugog reda regiona slike ograničenog maskom i definisan je kao $P(i,j|d,teta)$. Element (i,j) ove matrice predstavlja broj puta koliko se kombinaciju nivoa i i j javljaju u dva piksela na slici, koji su nalaze na rastojanju od d piksela pod uglom $teta$.

Tu spadaju: autokorelacija, korelacija, kontrast, združena energija, združena entropija...

Radiomics obeležja

Gray Level Size Zone (GLSZM). kvantifikuje zone nivoa sive na slici. Zona sivog nivoa se definiše kao broj povezanih vokselâ koji dele isti intenzitet sivog nivoa.

Tu spadaju: isticanje malih područja, isticanje velikih područja, nejednolikost sivih razina, varijacija sivih razina ...

Radiomics obeležja

Gray Level Run Length Matrix (GLRLM). kvantifikuje nizove sivog nivoa, koji su definisani kao dužina u broju piksela, uzastopnih piksela koji imaju istu vrednost nivoa sivog. U matrici dužine niza sivog nivoa $P(i,j|teta)$, element (i,j) opisuje broj prolazaka sa nivoom sive i i dužinom j koje se javljaju na slici (ROI) pod uglom $teta$.

Tu spadaju: isticanje kratkih sekvenci, isticanje dugih sekvenci, varijansa nivoa sive...



Klasifikacija

Na osnovu radiomics obeležja radimo klasifikaciju lezija na kancerogene i nekancerogene.



Podaci su izdijeljeni po pacijentima na trening, validacioni i test skup u odnosu 75:15:10 .



Kroz model propuštamo 4 varijante podataka:

- sve podatke (sve serije)
- samo serije pre ubacivanja kontrasta
- samo serije nakon ubacivanja kontrasta
- samo prvu seriju nakon ubacivanja kontrasta

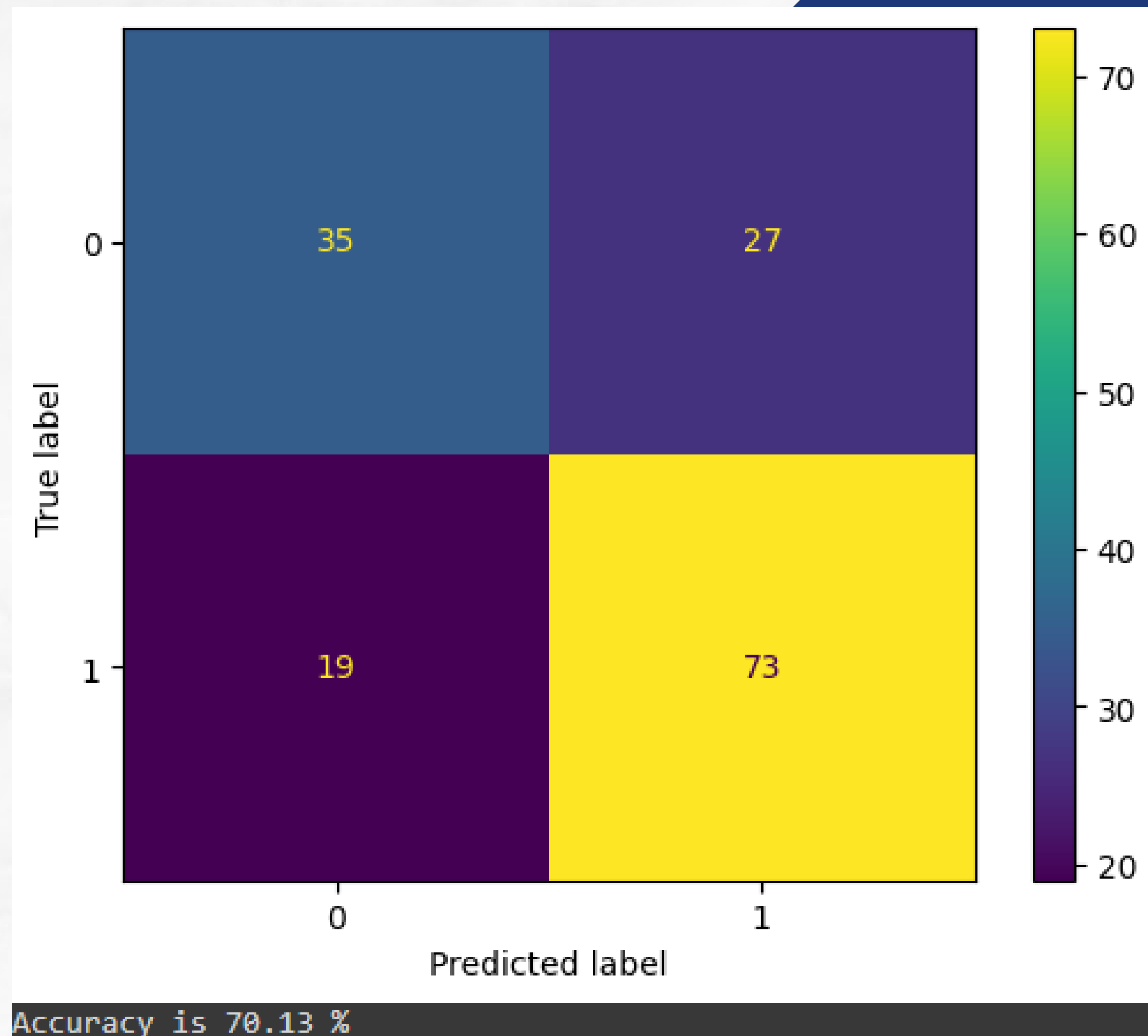


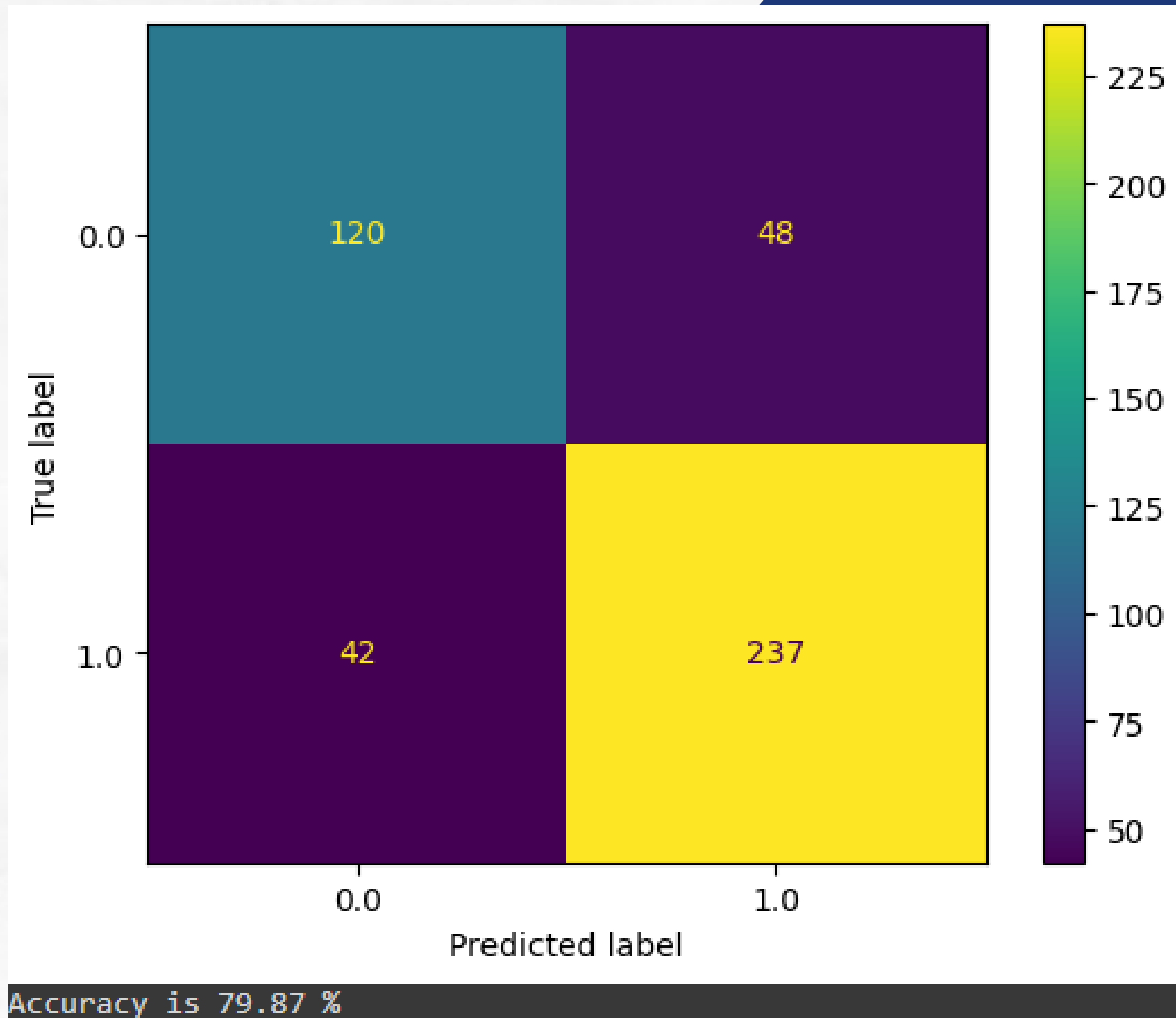
Put do najboljeg rezultata



kNN

daje najbolje rezultate za kombinaciju koja sadrži podatke samo iz 1. serije nakon puštanja kontrasta

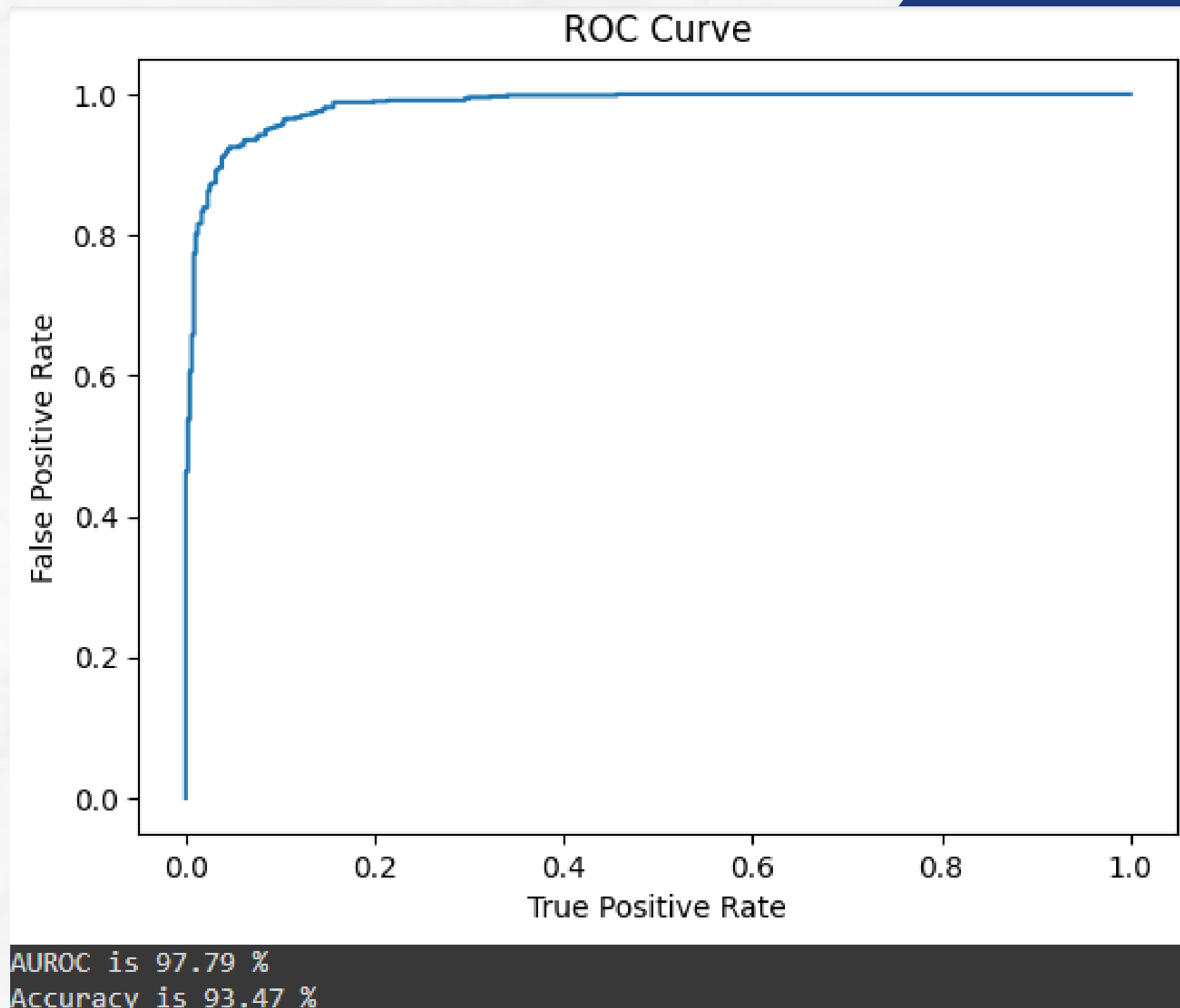




kNN

+ normalizacija

daje najbolje rezultate za kombinaciju koja sadrži podatke iz 3 serije nakon puštanja kontrasta



RandomForest

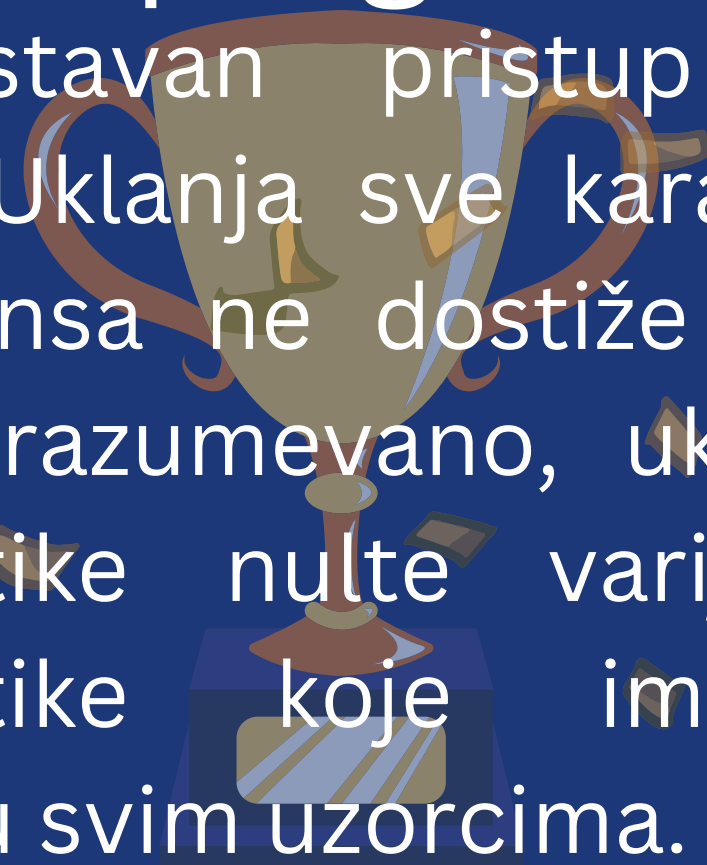
+ normalizacija

daje najbolje rezultate za
kombinaciju koja sadrži
podatke iz 1. serije nakon
puštanja kontrasta

Šta nam je još pomoglo da unapredimo rezultate? Selekcija obeležja!

Hi-kvadrat selektor je metoda filtera koja se koristi za kategoričke varijable i meri zavisnost između svakog obeležja i labele. Bira obeležja sa najvećom statistikom hi-kvadrata.

Selektor praga varijanse je jednostavan pristup odabiru obeležja. Uklanja sve karakteristike čija varijansa ne dostiže određeni prag. Podrazumevano, uklanja sve karakteristike nulte varijanse, tj. karakteristike koje imaju istu vrednost u svim uzorcima.



Selektovana

obeležja

od 79 obeležja,
samo 29 je
ocenjeno kao
korisno!

```
['original_firstorder_10Percentile' 'original_firstorder_90Percentile'  
'original_firstorder_Entropy' 'original_firstorder_InterquartileRange'  
'original_firstorder_Kurtosis' 'original_firstorder_Maximum'  
'original_firstorder_MeanAbsoluteDeviation' 'original_firstorder_Mean'  
'original_firstorder_Minimum'  
'original_firstorder_RobustMeanAbsoluteDeviation'  
'original_firstorder_RootMeanSquared' 'original_firstorder_Skewness'  
'original_firstorder_Uniformity' 'original_firstorder_Variance'  
'original_glcm_SumSquares' 'original_glszm_HighGrayLevelZoneEmphasis'  
'original_glszm_LargeAreaEmphasis'  
'original_glszm_LargeAreaHighGrayLevelEmphasis'  
'original_glszm_LowGrayLevelZoneEmphasis'  
'original_glszm_SmallAreaLowGrayLevelEmphasis'  
'original_glszm_ZonePercentage' 'original_glszm_ZoneVariance'  
'original_glrlm_GrayLevelVariance'  
'original_glrlm_HighGrayLevelRunEmphasis'  
'original_glrlm_LongRunEmphasis'  
'original_glrlm_LongRunHighGrayLevelEmphasis'  
'original_glrlm_LowGrayLevelRunEmphasis' 'original_glrlm_RunEntropy'  
'original_glrlm_RunPercentage']
```




Reference

Saha, A., Harowicz, M.R., Grimm, L.J., Kim, C.E., Ghate, S.V., Walsh, R. and Mazurowski, M.A., 2018. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. British journal of cancer, 119(4), pp.508-516. (A free version of this paper is available here: [PMC6134102](#))

[PyRadiomics](#)

[Feature Selection in PySpark: A Comprehensive Guide for Data Scientists](#) by Saturn Cloud



Hvala na pažnji!

