

Klasifikacija mladeža

Uvod i motivacija

Tema ovog projektnog zadatka je klasifikacija mladeža na benigne i maligne, analizom medicinskih slika. U okviru ovog projekta će se kao maligni mladeži analizirati samo melanomi. Dermatolozi posmatranjem i dermoskopijom na osnovu utvrđenih kriterijuma postavljaju radnu dijagnozu određenog tipa mladeža.

Postojanje široko dostupnog objektivnog sistema koji je u stanju da klasifikuje mladeže bi pružilo pomoć lekarima u dijagnostici i ranijem otkrivanju melanoma. Težnja takvog sistema ne bi bila da se nužno precizno razvrsta svaki mladež već da ne postoji maligni mladež koji nije prijavljen kao opasan.

Metodologija

Metoda kojom se u praksi vrši prepoznavanje tipa mladeža je ABCD(E) pravilo gde svako slovo ima odgovarajuće značenje.

A(*symmetry*) - sa smanjenjem simetrije mladeža raste rizik da je taj mladež melanom

B(*order*) - pravilnost ivica - benigni mladeži najčešće imaju ravnije ivice, dok su ivice melanoma nepravilnije, često sa usecima i reckama

C(*olor*) - boja - benigni mladeži su najčešće jednobojni sa malom varijacijom u nijansama, dok se kod melanoma često javlja više nijansi jedne ili čak više boja

D(*imension*)- dimenzije - melanom je najčešće veći od benignih mladeža

E(*volution*) - dodatna karakteristika koja se procenjuje na osnovu posmatranja promena mladeža u određenom vremenskom periodu

U ovom projektu akcenat je na prepoznavanju prve tri karakteristike s obzirom da ovaj set podataka nije sadržao podatke u kliničkim dimenzijama mladeža na slikama, kao ni neki drugi podatak na osnovu koga bi mogla da se odredi veličina mladeža, a postupak evolucije takođe nije bilo moguće pratiti.

Set podataka

Tokom realizacije projekta korišćen je javno dostupan set podataka HAM10000 koji sadrži nešto više od deset hiljada slika različitih vrsta mladeža od kojih je opisano nešto manje od osam hiljada.

U setu podataka postoji, pored slika običnih mladeža i melanoma, još pet dodatnih kategorija. Mladeži koji pripadaju tim kategorijama su zanemareni. Dodatno su zanemarene slike na kojima postoji vinjeta jer bi njihovo prisustvo skoro uvek dovelo do pogrešne segmentacije.

Zbog neizbalansiranosti podataka u približnom odnosu 5:1 u korist slika na kojima su obični mladeži uzorak je morao da bude dodatno smanjen i na kraju je analiza vršena na nešto više od 1000 slika svake klase.

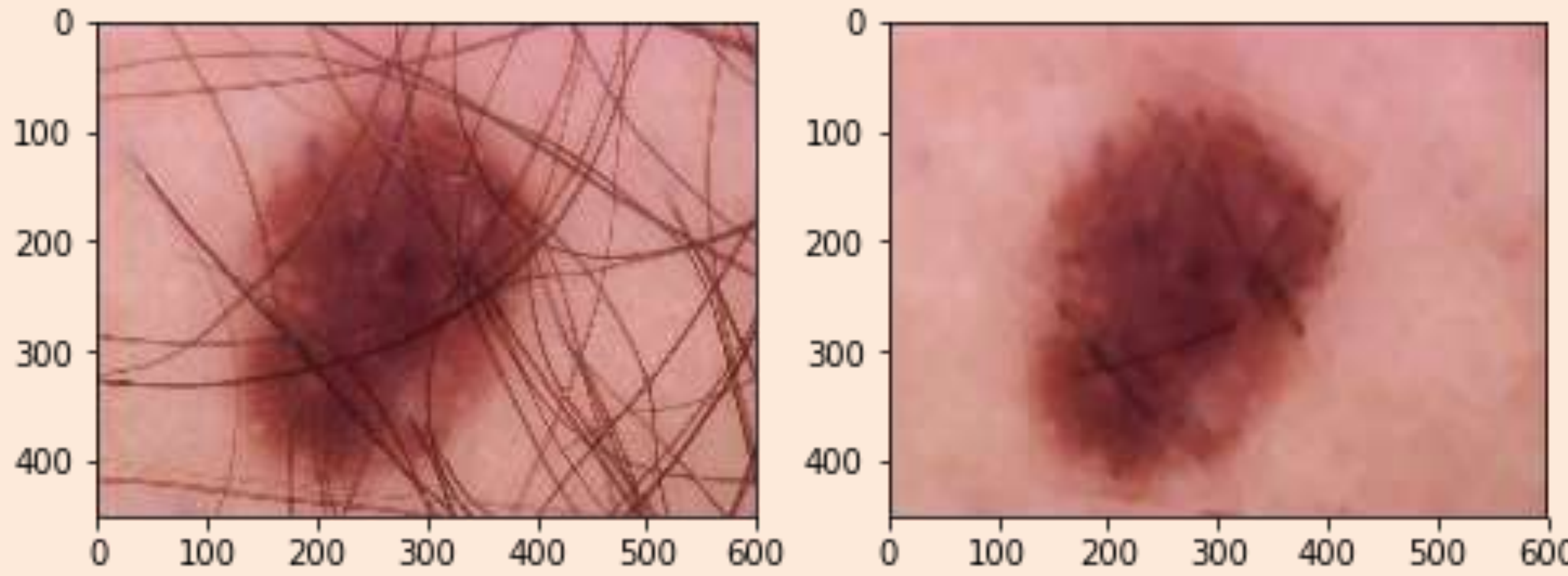
Obrada slike

Pretprocesiranje je prvi korak u obradi. Glavni cilj pretprocesiranja je bio da se ukloni šum, bolje izrazi kontura mladeža i uklone dlake koje postoje na slici. Za Uklanjanje dlaka korišćenjen je DullRazor algoritam koji se pokazao kao uspešan i u drugim israživanjima.

Posle pretprocesiranja vršena je segmentacija, pa ekstrakcija ciljanih obeležja. Da bi se odredio stepen simetrije prvo je bilo potrebno odrediti glavnu osu mladeža i to je učinjeno uz pomoć PCA (*Principal Component Analysis*) postupka. Nakon toga mladež je rotiran za 180 stepeni oko glavne ose i računata je površina na kojoj se originalni i zarotirani mladež poklapaju.

Pravilnost ivica je merena opisivanjem konveksnog poligona oko konture mladeža i poređenja površina konture mladeža i poligona.

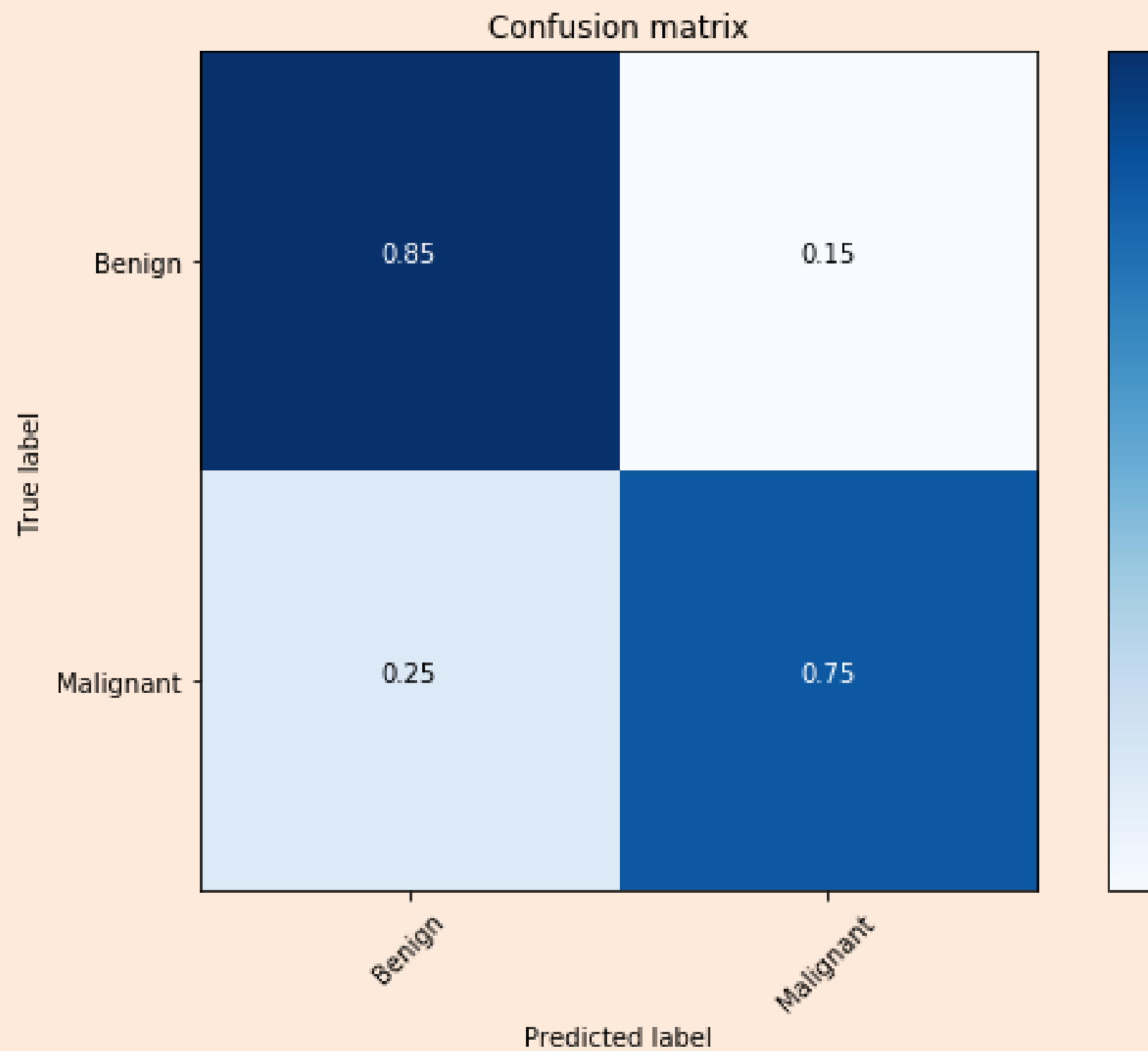
Boja mladeža je merena određivanjem standardne devijacije po sva tri kanala boje na piskelima koji pripadaju mladežu.



Klasifikacija

Treniranje modela i razvrstavanje mladeža u kategorije je rađeno na osnovu tri atributa koji su mereni i na osnovu opisa samih podataka. Za klasifikaciju su korišćene tehnike veštačke inteligencije: SVM (*Support Vector Machine*), k-NN (*k-nearest neighbors*) i MLP (*Multilayer perceptron*).

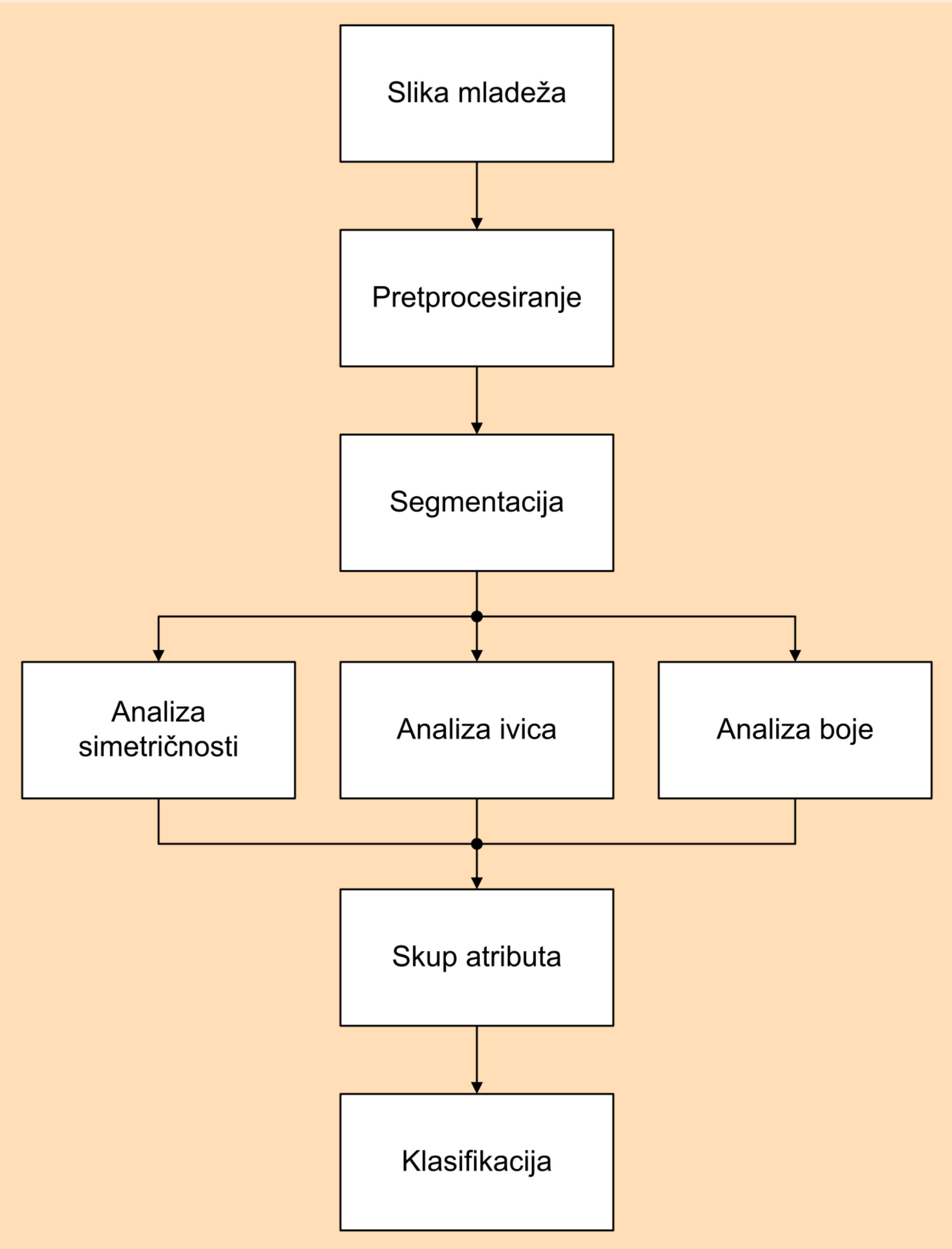
Tačnost SVM i k-NN-a je bila konstantna i najčešće veća od 75%. Optimalan broj suseda kod k-NN algoritma se kretao oko sedam. Za MLP bi morali dodatno da se podese parametri s obzirom da nije davao stabilne rezultate. Odziv modela je bio nešto slabiji u svim pristupima, a merene su i preciznost i F1-rezultat.



Zaključak i budući rad

Poboljšanja su moguća u raznim aspektima. Poboljšanje segmentacije i pretprocesiranja, kao i analiza grešaka sistema bi dovele do povećanja robustnosti rešenja.

Skup podaka može se proširiti novim klasama koje postoje u setu podataka. Uvođenje više niti u implementaciju ubrzalo bi proces analize slika.



Dijagram toka obrade slike za klasifikaciju mladeža