

# Klasifikacija raka kože korišćenjem konvolucijskih neuronskih mreža

Đorđe Jovanović, Jovan Tomić

Softversko inženjerstvo i infomacione tehnologije, Fakultet tehničkih nauka Novi Sad

## Motivacija

Rak kože predstavlja jednu od najčešćih vrsta kancera na svetu. Prvi znakovi ove bolesti mogu biti kvržice ili fleke nalik mladežima, koje bojom odudaraju od kože i razvijaju se mesecima, a ponekad i godinama. Najčešće se javljaju na delovima kože koji su izloženi suncu. Prepoznavanje ovih deformacija može dovesti do ranije dijagnoze, a samim tim i do povećanja šanse za uspešnim lečenjem. U prepoznavanju nam mogu pomoći konvolucijske neuronske mreže (CNN).

## Skup podataka

Korišćeni skup podataka sastoji se od 10 015 dermatoloških slika koje se mogu svrstati u sedam klasa raka kože. Tih sedam klasa su: *melanoma*, *melanocytic nevi*, *vascular lesions*, *dermatofibroma*, *benign keratosis-like lesions*, *basal cell carcinoma*, *Actinic keratoses and intraepithelial carcinoma*. Važno je napomenuti da skup podatak nije balansirani. Više od 60% skupa podatak predstavljaju slike koje pripadaju klasi *melanocytic nevi*.

## Pretprocesiranje podataka

Za rad nad ovim skupom podataka koristili smo dva modela koji zahtevaju različito pretprocesiranje podataka. Prvi model koristi slike čija je veličina smanjena na 32x32 i izvršena je normalizacija vrednosti matrice slike deljenjem piksela sa 255 (maksimalna vrednost piksela u RGB modelu). Za drugi model koristili smo već obrađene podatke koje smo dobili u sklopu skupa podataka. Slike su dimenzija 28x28 i takođe je izvršena normalizacija. Radi povećane preciznosti, podaci koje koristi drugi model su izbalansirani.

## Modeli

Kao što je već rečeno, koristili smo dva različita modela za klasifikaciju raka kože. U oba modela korišćene su konvolucijske neuronske mreže.

Prvi model se sastoji od sledećih slojeva:

1. Prvi konvolucijski blok
  - Conv2D(128, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2, 2))
  - Dropout(0.3)
2. Drugi konvolucijski blok
  - Conv2D(64, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2, 2))
  - Dropout(0.3)
3. Treći konvolucijski blok
  - Conv2D(32, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2, 2))
  - Dropout(0.3)
4. Flatten() sloj
5. Dense(16) i Dense(7, activation='softmax') sloj

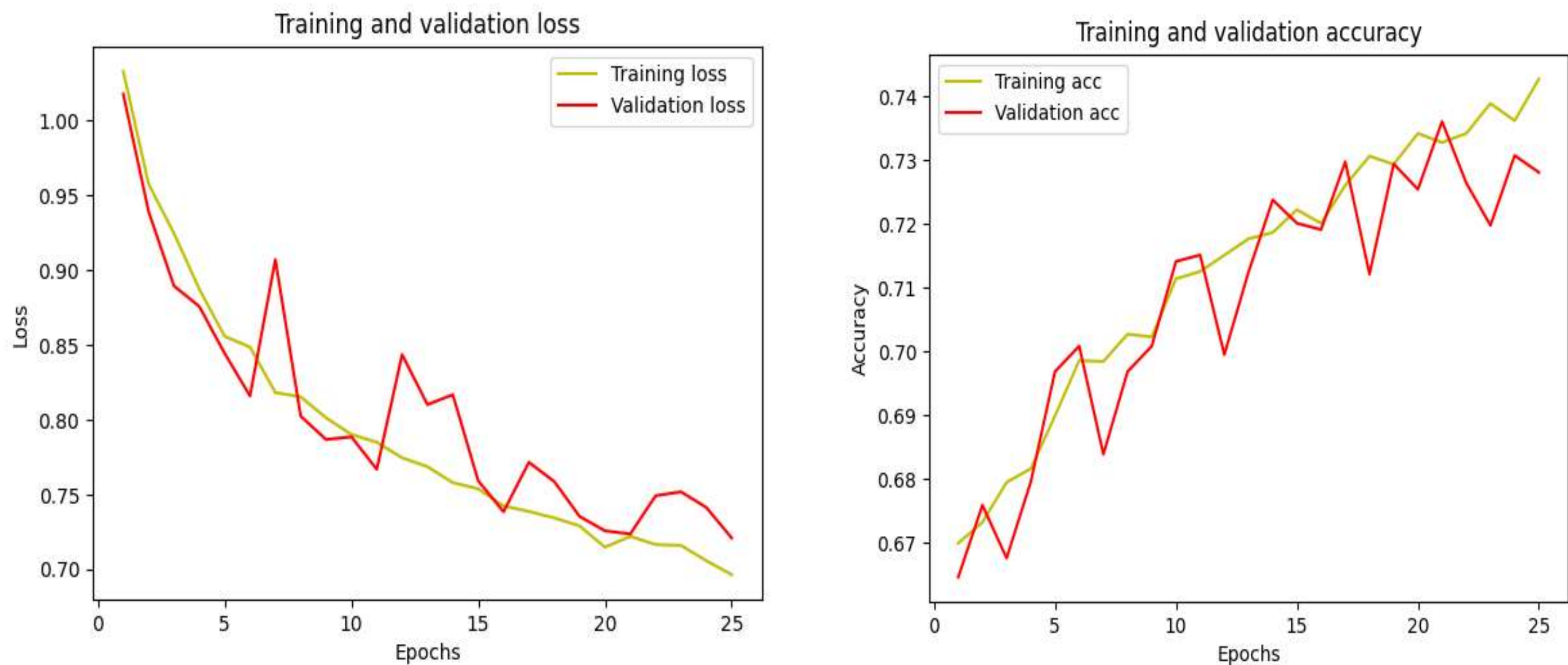
Drugi model se sastoji od sledećih slojeva:

1. Prvi konvolucijski blok
  - Conv2D(16, (3, 3), activation='relu')
  - Conv2D(32, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2,2))
2. Drugi konvolucijski blok
  - Conv2D(32, (3, 3), activation='relu')
  - Conv2D(64, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2,2))
3. Treći konvolucijski blok
  - Conv2D(64, (3, 3), activation='relu')
  - Conv2D(64, (3, 3), activation='relu')
  - MaxPooling2D(pool\_size=(2,2))

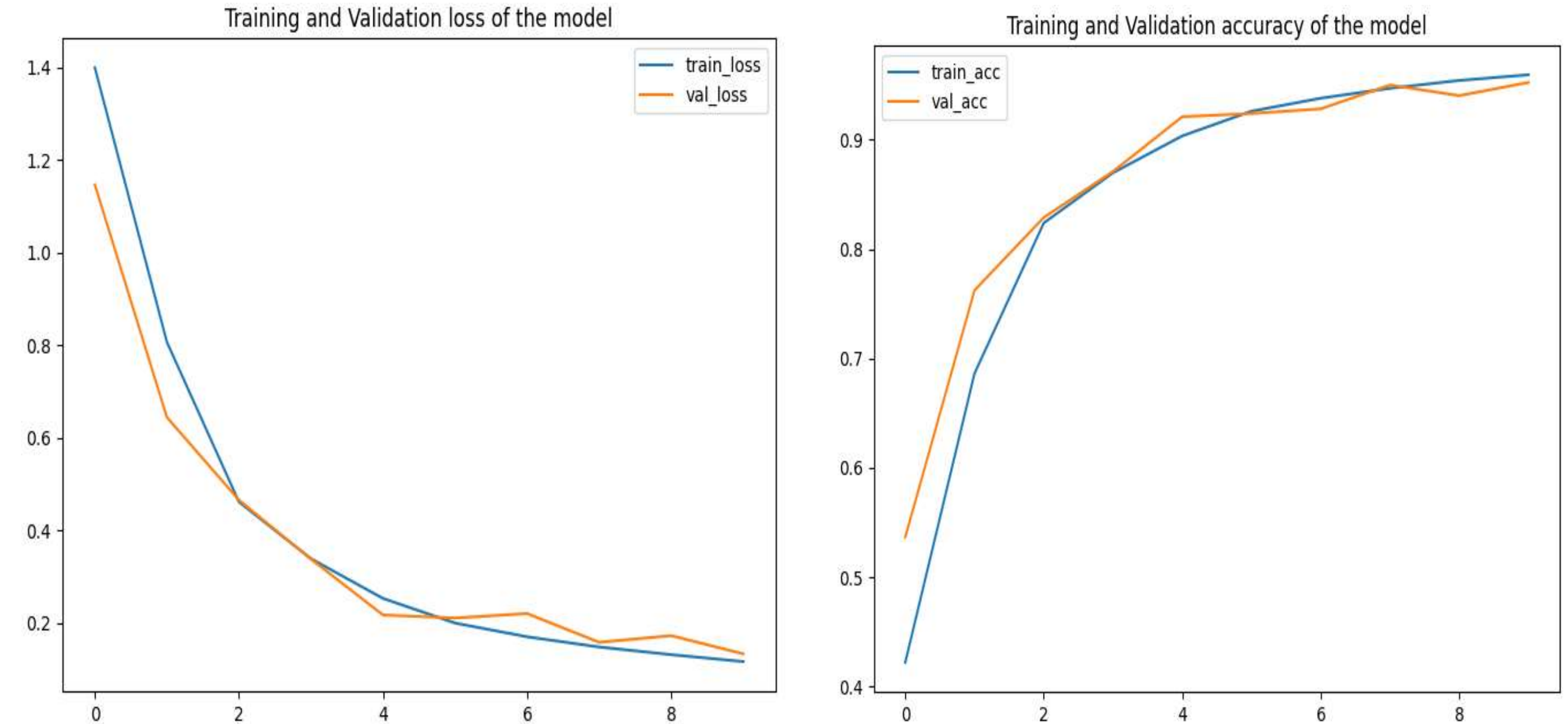
4. Flatten() sloj
5. Dva Dense(64) sloja
6. Dva Dense(32) sloja
7. Dense(7, activation='softmax')

## Rezultati

### Prvi model



### Drugi model



Tačnost prvog modela nakon 25 epoha je između 72-75%. Jedan od razloga slabijih rezultata ovog modela može biti nebalansirani ulazni podaci. Sa grafika iznad možemo uočiti da preciznost prilikom validacije nije stabilna. Ova pojava se može javiti kao rezultat više različitih stvari. Jedan od mogućih uzroka jeste *Overfitting*. Drugi model u 10 epoha dostigne preciznost između 92-96%. Ovaj model, za razliku od prvog, koristi izbalansirani skup podataka.