

# Izveštaj

## Uvod

Tumor, odnosno neoplazma ili novotvorina, označava skup promenjenih ćelija koje pokazuju nepravilan i progresivan rast. Tumorske bolesti imaju ozbiljnu prognozu, a neke su opasne po život, te su danas aktuelna tema traženja novih načina lečenja.

Nastaje kao posledica unutrašnjih i spoljašnjih faktora koji deluju na organizam. Može se definisati i kao abnormalna nakupina tkiva čiji je rast nesvrsishodan, autonoman i nadmašuje rast normalnog tkiva, te nastavlja da raste i nakon prestanka uzroka, koji je praktično nepoznat.

Novonastale ćelije se od polaznih ćelija razlikuju strukturno i funkcionalno, ali se u većini slučajeva može prepoznati od koje vrste ćelija ili tkiva je nastao tumor. U osnovi, novotvorine nastaju kada se izgubi regulacija kontrolnih mehanizama rasta ćelija. Glavna podela je na maligne (zloćudne) i benigne (dobroćudne).

## Problem

U ovom projektu je implementirana konvoluciona neuronska mreža (CNN) za klasifikaciju četiri različite vrste tumora na snimcima magnetne rezonance mozga (MRI snimcima).

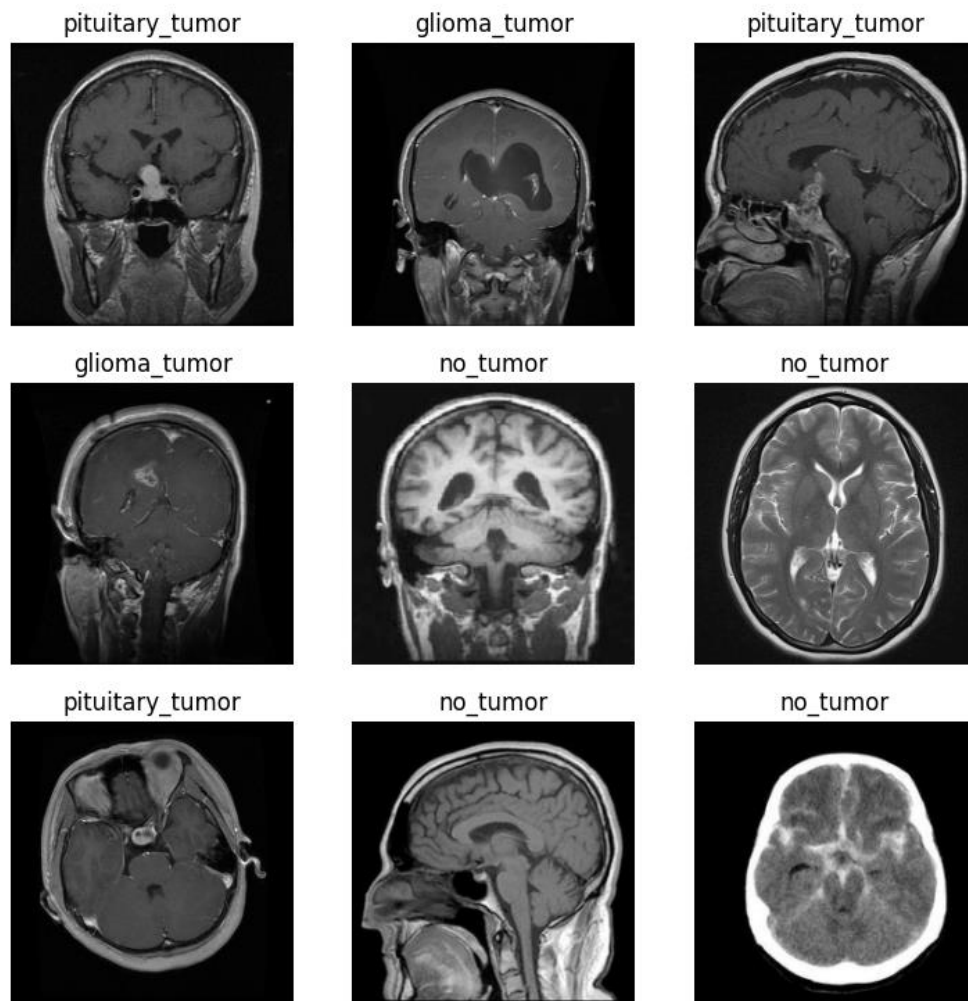
Postoje više različitih vrsta tumora, u ovom projektu klasifikaciju vršimo na četiri kategorije: bez tumora, gliom, tumor hipofize, meningiom

Sam skup podatak ćemo izdeliti da delove za učenje, validaciju i testiranje. Skup podataka za validaciju i treniranje neuronske mreže sastoji se iz 2870 slika, podeljenih principom 80/20 (80% podataka se nalazi u skupu za treniranje, a 20% u skupu za validiranje).

Preostale 394 slike koriste se za testiranje i evaluiranje same mreže, na osnovu čijih rezultata izvodimo zaključke.

## Prikaz originalnih podataka podataka

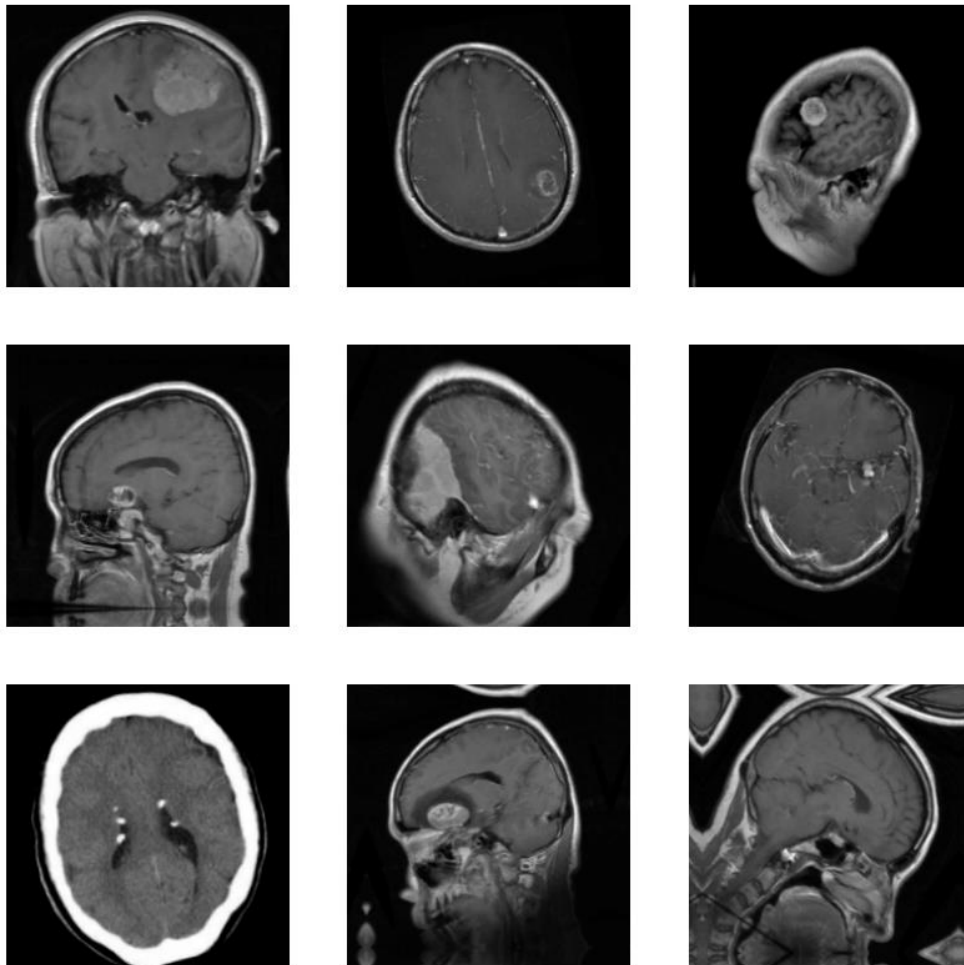
Slike su učitane u “grayscale” modu boja kako bismo smanjili broj korišćenih kanala boja radi optimizacije (naše slike su svakako izmedju crne, bele i sive). Koristimo “Shuffle” mod kako bismo na slučajan način odabrali slike.



## Prikaz izmenjenih podataka

Data Augmentation – proces u kome vršimo razne transformacije slika, kako bismo povećali skup za trening i smanjili mogućnost overfitting-a.

Korišćene metode: RandomFlip, RandomBrightness, RandomZoom, RandomRotation



## Pojašnjenje rezultata

Matrica konfuzije služi nam kako bismo evaluirali model i uočili gde model greši. Brojevi u ćelijama glavne dijagonale predstavljaju brojeve tačno predpostavljenih tumora, dok brojevi u ćelijama iznad i ispod glavne dijagonale predstavljaju promašaje.

## Rešenje korišćenjem sopstvene arhitekture modela

### Konvoluciona neuronska mreža

Konvoluciona neuronska mreža se sastoji od ulaznog sloja, skrivenih slojeva i izlaznog sloja. U konvolucionoj neuronskoj mreži, skriveni slojevi uključuju jedan ili više slojeva koji izvode konvolucije.

Korišćeni slojevi:

Conv2D, BatchNormalization, MaxPooling2D, Flatten, Dense, Dropout, Dense

Za aktivacionu funkciju korišćena je ReLU funkcija.

Za loss funkciju smo koristili categorical crossentropy.

Kod hiperparametara, za veličinu kernela smo uzeli 3x3, stride 1, padding same (0)

Za “pooling” operaciju korišćena je tehnika “max pooling” sa veličinom 2x2.

## Rešenje korišćenjem nekih od “pretrained” modela

Skup podataka je previše mali da bi se CNN obučio od početka i dao pozitivne rezultate. Zbog ovog problema pokušaćemo drugi pristup, takozvani prenos učenja - Transfer Learning. Prenos učenja kao opšti pojam se odnosi na ponovnu upotrebu znanja jednog zadatka na drugi. Posbeno za konvolucione neuronske mreže, mnoge karakteristike slike su zajedničke za različite skupove podataka (npr. linije, ivice). Iz tog razloga se cnn veoma retko obučava potpuno od nule jer je teško doći do velikih skupova podataka i velikih računarskih resursa.

U ovom projektu korišćena su 2 modela, DenseNet169 i Efficientb0 kao “pretrained” modeli. U nastavku će biti prikazani rezultati. Za više detalja i strukturu samog modela posetiti linkove u prilogu.

## EfficientNetB0 (imagenet) i DenseNet169

EfficientNet: <https://paperswithcode.com/method/efficientnet>

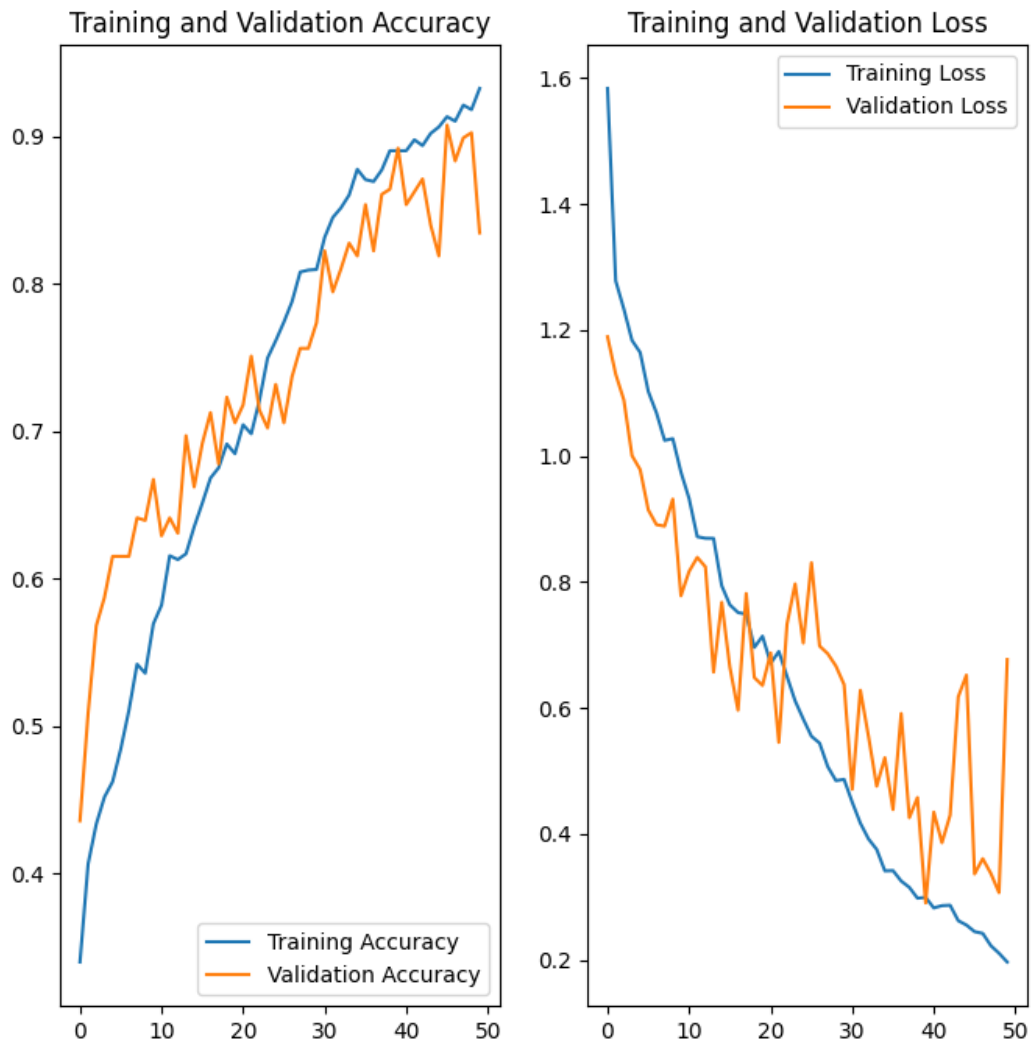
DenseNet: <https://paperswithcode.com/method/densenet>

## Zaključak

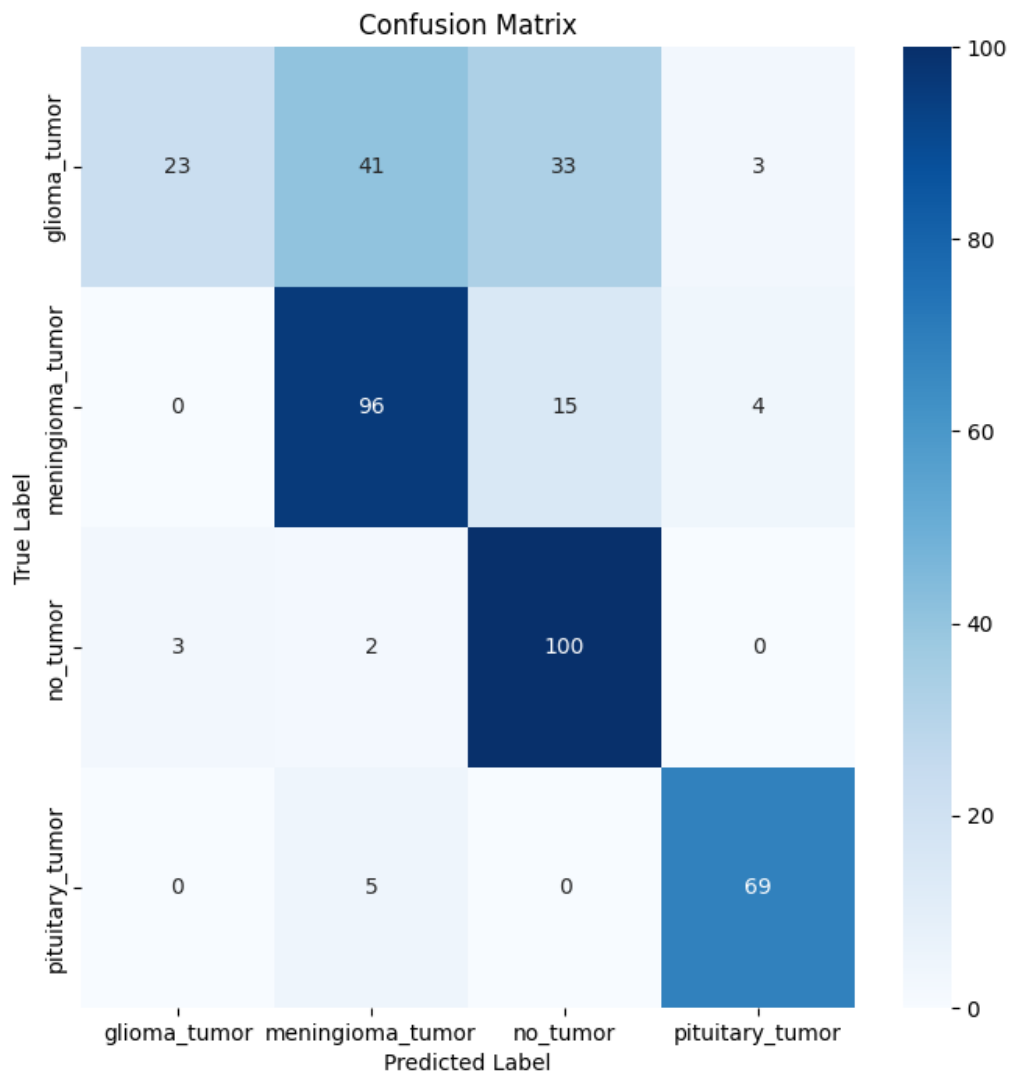
Veliki problem nastaje u maloj raznolikosti slika, iako 3000 slika zvuči kao veliki skup za ovaj problem je potrebno više slika. Takođe, jedan od problema je to što se doktori pri donošenju dijagnoze ne oslanjaju isključivo na slike, već i na neke dodatne podatke i zaključke, koji u ovom projektu nisu obuhvaćeni.

## Prikaz rezultata

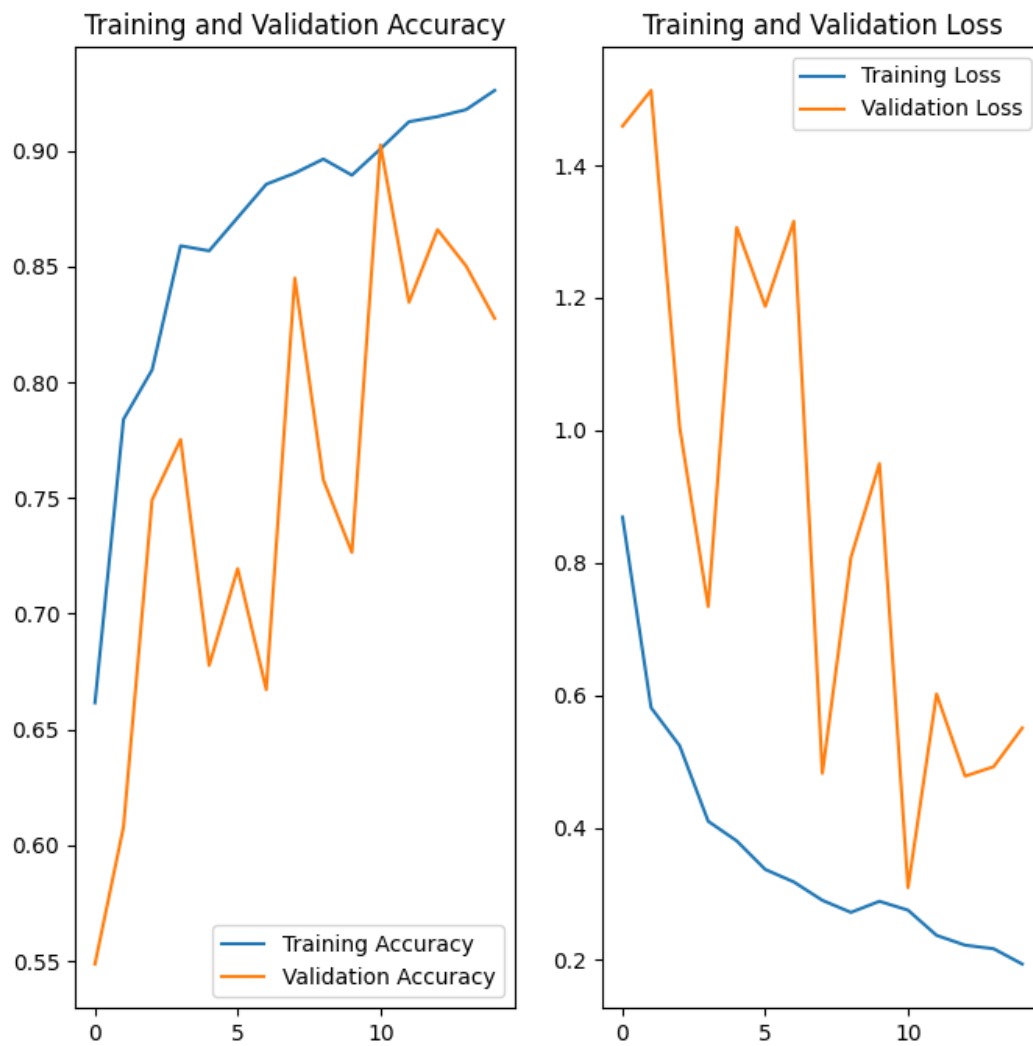
Parameters / Model	Custom model	EfficientNetB0	DenseNet169
Accuracy	<b>0.73</b>	<b>0.73</b>	<b>0.61</b>



Slika 1 Treniranje modela (custom model)

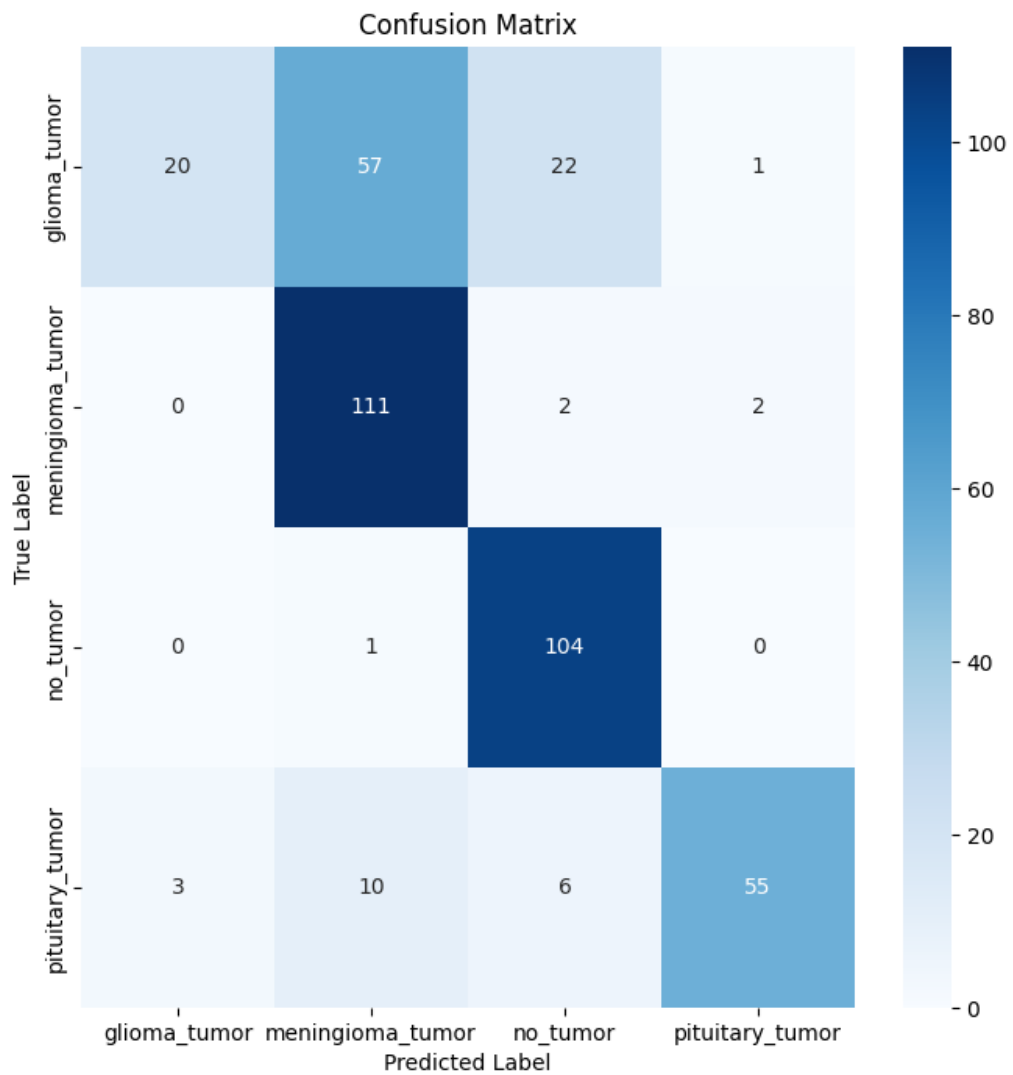


Slika 2 matrica konfuzije za model (custom model)

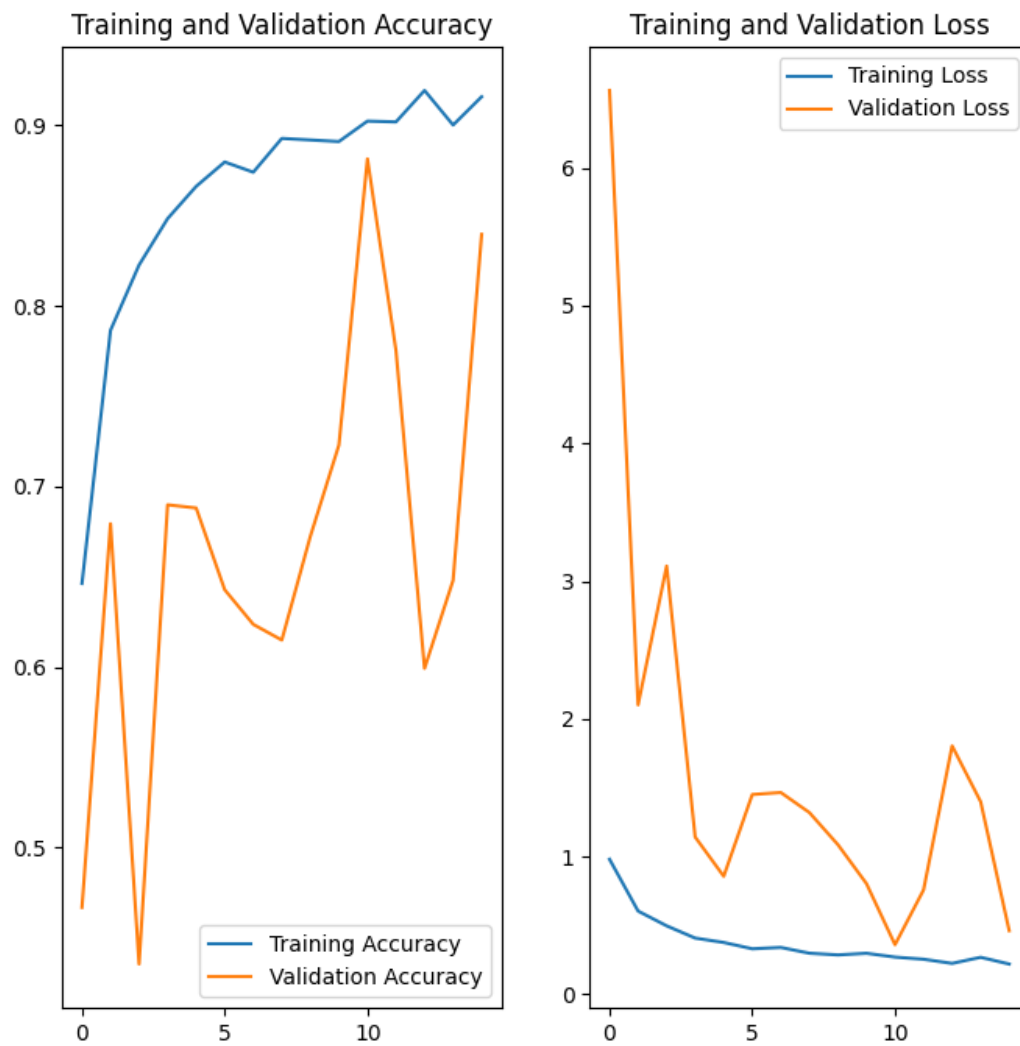


Slika 3 EfficientNetB0 treniranje

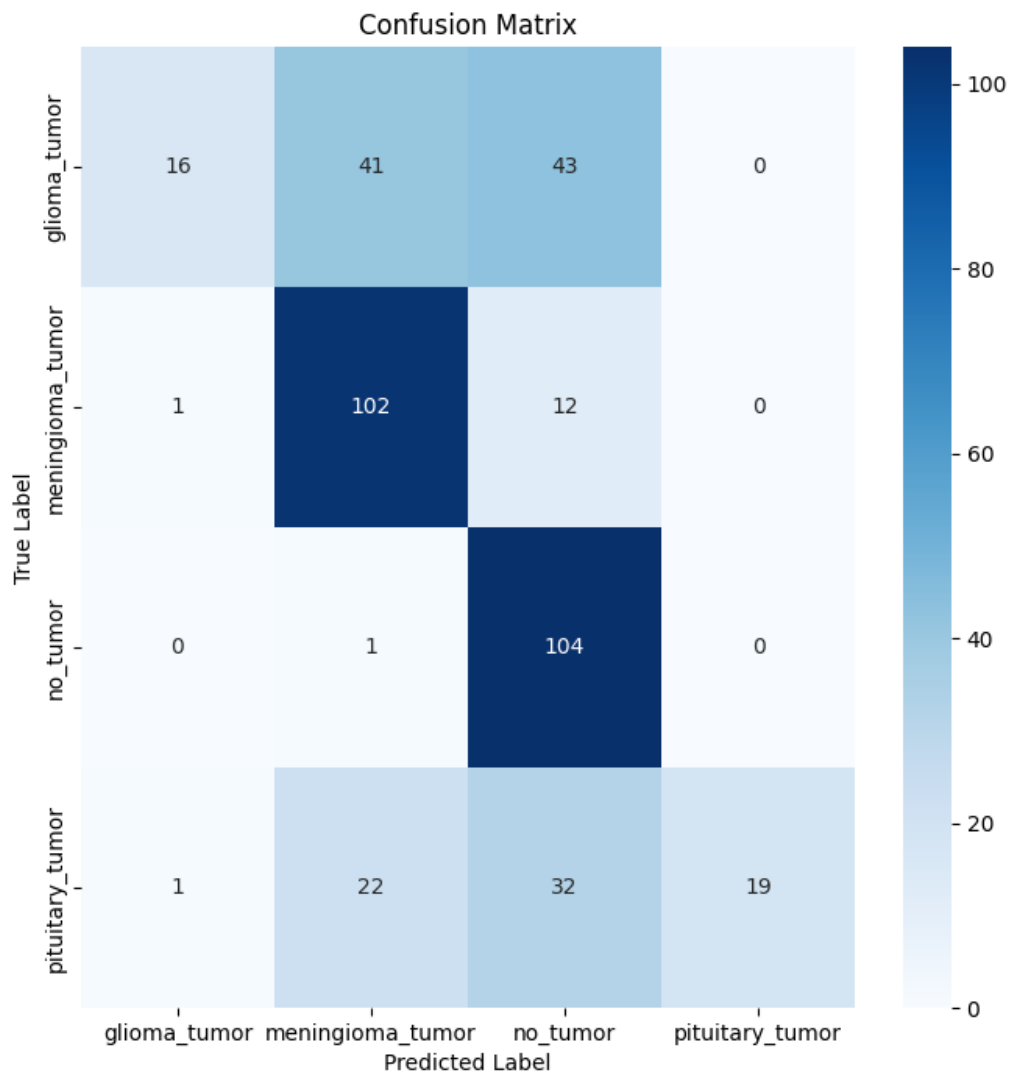




Slika 4 EfficientNetB0 matrica konfuzije



Slika 5 DenseNet169 obuka modela



Slika 6 matrica konfuzije za DenseNet169