

KLASYFIKACJA OBRAZÓW GUZÓW MÓZGU
PROJEKT ZALICZENIOWY WUM 2022/23

Urszula Baranowska (s26544)

PJATK

10.01.2023

KLASYFIKACJA OBRAZÓW GUZÓW MÓZGU

- ▶ Celem mojego projektu jest zbudowanie prostego klasyfikatora obrazów guzów mózgu.
- ▶ Wykorzystam zbiór danych autorstwa Jun Cheng brain tumor dataset dostępny online:
<https://doi.org/10.6084/m9.figshare.1512427.v5>
- ▶ Ten zbiór danych został przez autorów wykorzystany w dwóch publikacjach dotyczących bardziej zaawansowanych metod klasyfikacji guzów mózgu:
 - <https://doi.org/10.1371/journal.pone.0140381>
 - <https://doi.org/10.1371/journal.pone.0157112>
- ▶ Zbiór zawiera 3064 obrazy (MRI) od 233 pacjentów, 3 rodzaje guzów (guzy przysadki, oponiaki i glejaki). Obrazy są w formacie .mat (matlab), ale można je przekształcić na np.array o wymiarach 512x512, każdy obraz ma dodatkowo ID pacjenta, etykietę (1-3), informacje o położeniu guza na obrazie (współrzędne granic i maska)
- ▶ 15 obrazów ma inne wymiary i dla uproszczenia zostaną wykluczone z analizy, zatem ostatecznie zostanie 3049 obrazów: 1429 glejaków, 915 guzów przysadki i 708 oponiaków
- ▶ Pierwszym krokiem będzie zbudowanie prostego klasyfikatora korzystając z pakietu sklearn - SVM (baseline), następnie będą podejmowane próby zbudowania sieci neuronowej (najlepiej konwolucyjnej) i stosowania regularyzacji zależnie od mocy obliczeniowych. Dla uproszczenia maska i położenie guza zostaną na razie pominięte

PRZYKŁADOWE OBRAZY (Z NAŁOŻONĄ MASKĄ)

image 252, patient ID: 494949485553, tumor type: pituitary tumor

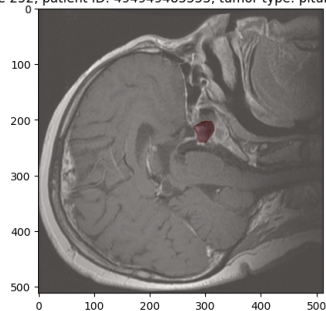


image 1234, patient ID: 494848515448, tumor type: meningioma

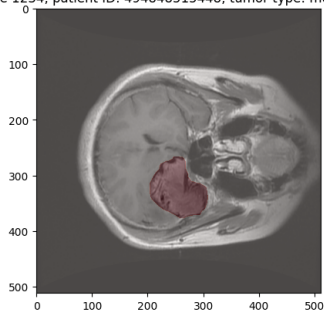
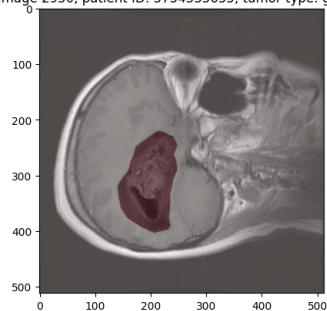


image 2950, patient ID: 5754535655, tumor type: glioma

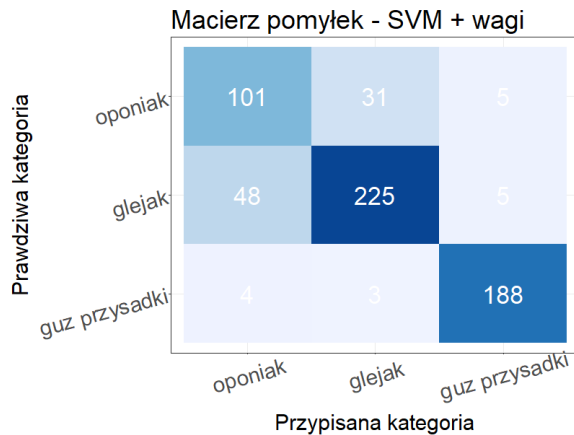
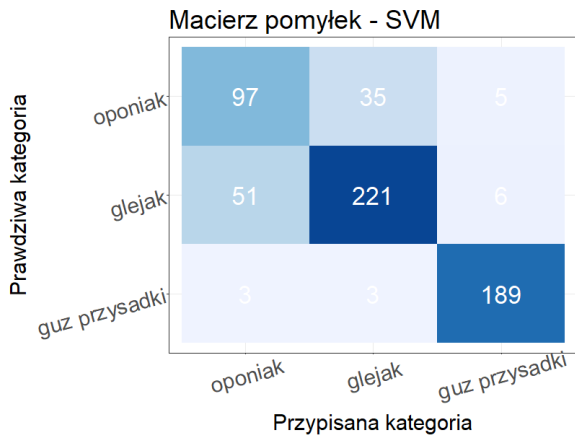


PRZETWARZANIE DANYCH

- ▶ Napisanie 4 funkcji ułatwiających wczytanie danych z formatu .mat do np.array, zamianę etykiet z 1-3 na 0-2 oraz nazwy guzów, łatwe wyświetlenie obrazów z podpisem
- ▶ Sprawdzenie zawartości zbioru (ile obrazów, wymiary, rozkład klas) - te informacje znajdują się na wcześniejszym slajdzie
- ▶ Normalizacja min-max - w oryginalnym zbiorze wartości od 0 do prawie 13000
- ▶ W razie potrzeby - zmiana kształtu (np.reshape) lub zmiana etykiet na one-hot-encoding (tf.keras.utils.to_categorical)
- ▶ Podział na zbiór treningowy (60%), walidacyjny (20%) i testowy (20%) (sklearn.model_selection.train_test_split)

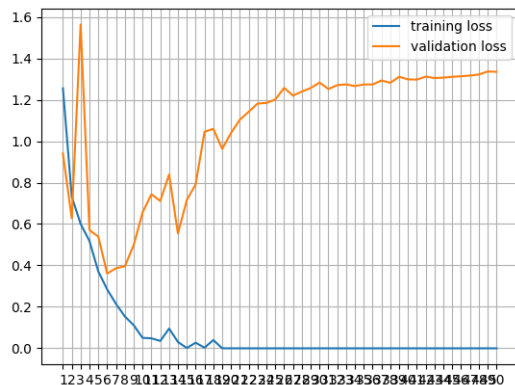
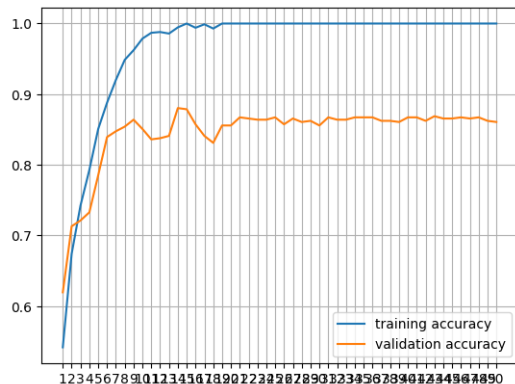
SVM - SKLEARN.SVC.LINEARSVC

- Dokładność - 83.11%, dodanie wag polepszyło wynik do 84.26%



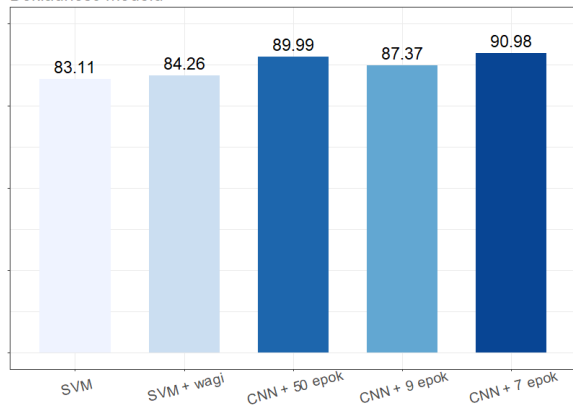
SIEĆ KONWOLUCYJNA

- ▶ warstwa wejściowa: kształt - (512, 512, 1)
- ▶ warstwa konwolucyjna: filtry - 32, rozmiar filtra - 8, funkcja aktywacji - ReLU
- ▶ warstwa max pooling - rozmiar 2
- ▶ warstwa konwolucyjna: filtry - 64, rozmiar filtra - 8, funkcja aktywacji - ReLU
- ▶ warstwa max pooling - rozmiar 2
- ▶ warstwa konwolucyjna: filtry - 128, rozmiar filtra - 4, funkcja aktywacji - ReLU
- ▶ warstwa flatten
- ▶ warstwa wyjściowa: 3 neurony, funkcja aktywacji - softmax
- ▶ funkcja straty: categorical cross entropy
- ▶ liczba epok: 50 (testowy i walidacyjny), potem 9 i 7 (cały zbiór testowy)

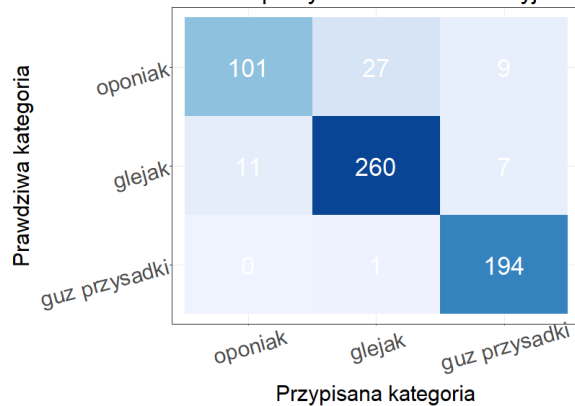


WYNIKI

Dokładność modelu



Macierz pomyłek - sieć konwolucyjna



PRZYKŁADY BŁĘDNIIE SKLASYFIKOWANYCH OBRAZÓW)

