

Klasyfikacja guzów mózgu na obrazach MRI

Krzysztof Deka, Jakub Olejniczak

5 stycznia 2026

1 Wstęp

Obrazowanie metodą rezonansu magnetycznego (MRI) odgrywa kluczową rolę w diagnostyce chorób mózgu, w szczególności w wykrywaniu oraz klasyfikacji nowotworów. Analiza obrazów MRI jest procesem czasochłonnym i wymaga dużego doświadczenia ze strony specjalistów, co sprawia, że coraz częściej poszukuje się automatycznych metod wspomagających diagnostykę medyczną. Jednym z najdynamiczniej rozwijających się obszarów w tym zakresie są techniki uczenia maszynowego i głębokiego uczenia, które wykazują wysoką skuteczność w zadaniach analizy obrazów.

Przez wiele lat dominującym podejściem w klasyfikacji obrazów były konwolucyjne sieci neuronowe (CNN), które dzięki lokalnemu przetwarzaniu informacji oraz hierarchicznej ekstrakcji cech znalazły szerokie zastosowanie w analizie danych medycznych. W ostatnich latach coraz większą popularność zyskują jednak architektury transformerowe, pierwotnie opracowane na potrzeby przetwarzania języka naturalnego, a następnie zaadaptowane do analizy obrazów. Modele te, oparte na mechanizmie samo-uwagi, umożliwiają globalne modelowanie zależności pomiędzy fragmentami obrazu i często osiągają bardzo wysoką skuteczność klasyfikacji.

Celem niniejszego sprawozdania jest porównanie skuteczności klasycznej konwolucyjnej sieci neuronowej VGG19 oraz nowoczesnej architektury transformerowej Swin Transformer w zadaniu klasyfikacji obrazów MRI mózgu. Porównanie przeprowadzono na wspólnym zbiorze danych obejmującym cztery klasy: glejak, oponiak, gruczolak przysadki oraz obrazy bez obecności nowotworu. W ramach pracy oceniono jakość klasyfikacji z wykorzystaniem standardowych metryk, takich jak dokładność (accuracy), precyzja, czułość oraz miara F1, a także przeanalizowano macierze pomyłek oraz przebieg procesu uczenia modeli.

2 Opis zbioru danych

W eksperymentach wykorzystano publicznie dostępny zbiór danych obrazów rezonansu magnetycznego mózgu (MRI Brain Tumor Dataset), który jest często stosowany w badaniach nad automatyczną detekcją i klasyfikacją nowotworów

mózgu. Zbiór ten zawiera obrazy MRI przedstawiające zarówno przypadki patologiczne, jak i obrazy zdrowe, co umożliwia ocenę skuteczności modeli w zadaniu wieloklasowej klasyfikacji.

2.1 Klasy i struktura danych

Zbiór danych obejmuje cztery klasy:

- **glioma** – obrazy przedstawiające glejaki mózgu,
- **meningioma** – obrazy przedstawiające oponiaki,
- **pituitary** – obrazy przedstawiające guzy przysadki mózgowej,
- **notumor** – obrazy bez obecności nowotworu.

Dane zostały zorganizowane w strukturze katalogów odpowiadającej poszczególnym klasom, co umożliwiło ich bezpośrednie wykorzystanie w bibliotekach do uczenia głębokiego. Całkowita liczba obrazów w zbiorze wynosiła **7023** próbek.

2.2 Podział zbioru danych

Zbiór danych został podzielony na trzy rozłączne podzbiory:

- zbiór treningowy – **4571** obrazów,
- zbiór walidacyjny – **1141** obrazów,
- zbiór testowy – **1311** obrazów.

Podział danych został przeprowadzony w sposób zapewniający zachowanie reprezentatywności wszystkich klas w każdym z podzbiorów. Zbiór testowy nie był wykorzystywany na żadnym etapie treningu modeli i posłużył wyłącznie do końcowej oceny ich skuteczności.

2.3 Przygotowanie danych

Przed przekazaniem obrazów do modeli wszystkie próbki zostały przeskalowane do rozmiaru 224×224 piksele, co było zgodne z wymaganiami zastosowanych architektur sieci neuronowych. Dodatkowo obrazy poddano normalizacji wartości pikseli. Takie przygotowanie danych pozwoliło na ujednolicenie wejścia do modeli oraz stabilizację procesu uczenia.

3 Opis architektur modeli

W niniejszym rozdziale przedstawiono architektury dwóch modeli uczenia głębokiego wykorzystanych w eksperymentach: klasycznej konwolucyjnej sieci neuronowej VGG19 oraz nowoczesnej architektury transformerowej Swin Transformer. Modele te reprezentują dwa odmienne podejścia do analizy obrazów, co umożliwia ich bezpośrednie porównanie w dalszej części pracy.

3.1 Konwolucyjna sieć neuronowa VGG19

VGG19 jest głęboką konwolucyjną siecią neuronową zaproponowaną przez Simonyana i Zissermana, charakteryzującą się prostą i regularną strukturą. Architektura ta składa się z 19 warstw uczących się, w tym 16 warstw konwolucyjnych oraz 3 warstw w pełni połączonych. Podstawowym elementem sieci są bloki konwolucyjne wykorzystujące filtry o rozmiarze 3×3 , po których następują warstwy agregujące typu max-pooling.

Hierarchiczna struktura VGG19 umożliwia stopniową ekstrakcję cech obrazu – od prostych krawędzi i tekstur w początkowych warstwach, po bardziej złożone reprezentacje w głębszych warstwach sieci. Dzięki temu model ten znalazł szerokie zastosowanie w zadaniach klasyfikacji obrazów, w tym w analizie danych medycznych.

W przeprowadzonych eksperymentach wykorzystano wersję VGG19 wstępnie wytrenowaną na zbiorze ImageNet. Warstwy konwolucyjne pełniły rolę ekstraktora cech, natomiast końcowa część sieci została dostosowana do zadania czteroklasowej klasyfikacji obrazów MRI mózgu poprzez dodanie nowej warstwy klasyfikacyjnej.

3.2 Architektura transformerowa Swin Transformer

Swin Transformer jest nowoczesną architekturą opartą na mechanizmie samo-uwagi (self-attention), zaprojektowaną z myślą o efektywnej analizie obrazów. W przeciwieństwie do klasycznych Vision Transformerów, Swin Transformer wykorzystuje mechanizm uwagi obliczanej w lokalnych, przesuwanych oknach (ang. *shifted windows*), co znacząco redukuje złożoność obliczeniową i umożliwia skalowanie modelu do obrazów o większej rozdzielczości.

Architektura Swin Transformer przetwarza obraz poprzez podział na niewielkie fragmenty (patche), które następnie są analizowane za pomocą wielogłowej samo-uwagi. Przesuwanie okien pomiędzy kolejnymi warstwami pozwala na stopniowe modelowanie zależności globalnych, przy jednoczesnym zachowaniu efektywności obliczeniowej. Dzięki temu model łączy zalety lokalnej analizy znanej z sieci konwolucyjnych z globalnym kontekstem charakterystycznym dla transformerów.

W niniejszej pracy zastosowano wariant Swin Transformer Tiny, wstępnie wytrenowany na zbiorze ImageNet, który następnie poddano procesowi dostrajania (fine-tuningu) do zadania klasyfikacji obrazów MRI mózgu. Ostatnia warstwa klasyfikacyjna została dostosowana do liczby klas występujących w analizowanym zbiorze danych.

4 Metodyka badań

W niniejszym rozdziale przedstawiono metodykę przeprowadzonych eksperymentów, obejmującą sposób przygotowania danych, konfigurację procesu uczenia modeli oraz zastosowane metryki oceny jakości klasyfikacji. W celu zapew-

nienia rzetelnego porównania obu architektur zastosowano ten sam zbiór danych testowych oraz identyczny zestaw miar ewaluacyjnych.

4.1 Przygotowanie danych

Obrazy MRI zostały poddane wstępnemu przetwarzaniu przed przekazaniem ich do modeli uczenia głębokiego. Wszystkie próbki przeskalowano do rozmiaru 224×224 piksele, co było zgodne z wymaganiami architektur VGG19 oraz Swin Transformer. Dodatkowo zastosowano normalizację wartości pikseli w celu ujednolicenia zakresu danych wejściowych oraz poprawy stabilności procesu uczenia.

Zbiór danych został podzielony na część treningową, walidacyjną oraz testową. Zbiór walidacyjny służył do monitorowania procesu uczenia i doboru parametrów, natomiast zbiór testowy był wykorzystywany wyłącznie do końcowej oceny jakości wytrenowanych modeli.

4.2 Proces uczenia modelu VGG19

Model VGG19 został zaimplementowany z wykorzystaniem biblioteki TensorFlow oraz interfejsu Keras. W eksperymentach zastosowano wersję sieci wstępnie wytrenowaną na zbiorze ImageNet, co umożliwiło wykorzystanie wcześniej wyuczonych reprezentacji wizualnych. Warstwy konwolucyjne pełniły rolę ekstraktora cech, natomiast końcowa część sieci została dostosowana do zadania czteroklasowej klasyfikacji obrazów MRI.

Proces uczenia modelu przeprowadzono przez **20 epok**. Jako funkcję straty zastosowano entropię krzyżową, natomiast optymalizację wag realizowano przy użyciu algorytmu Adam. W trakcie treningu monitorowano wartości funkcji straty oraz dokładności klasyfikacji dla zbioru treningowego i walidacyjnego, co umożliwiło ocenę stabilności procesu uczenia.

4.3 Proces uczenia modelu Swin Transformer

Model transformerowy Swin Transformer został zaimplementowany z wykorzystaniem biblioteki Hugging Face Transformers oraz frameworka PyTorch. Wykorzystano wariant Swin Transformer Tiny wstępnie wytrenowany na zbiorze ImageNet. Ze względu na różnicę liczby klas pomiędzy zadaniem źródłowym a analizowanym problemem, końcowa warstwa klasyfikacyjna modelu została zainicjalizowana ponownie i dostosowana do czterech klas.

Proces dostrajania (fine-tuning) modelu przeprowadzono przez **5 epok**. Podobnie jak w przypadku modelu VGG19 zastosowano funkcję straty opartą na entropii krzyżowej oraz optymalizator Adam. W trakcie treningu rejestrowano wartości funkcji straty oraz dokładności klasyfikacji dla zbioru treningowego, co umożliwiło analizę przebiegu procesu uczenia.

4.4 Metryki oceny jakości

Jakość klasyfikacji obu modeli oceniono z wykorzystaniem standardowych metryk stosowanych w zadaniach wieloklasowej klasyfikacji obrazów. W szczególności obliczono:

- dokładność klasyfikacji (accuracy),
- precyzję (precision),
- czułość (recall),
- miarę F1 (F1-score) w ujęciu makro.

Dodatkowo dla obu modeli wygenerowano macierze pomyłek (confusion matrix), które umożliwiły szczegółową analizę błędów klasyfikacji dla poszczególnych klas. Takie podejście pozwoliło na kompleksową ocenę skuteczności badanych architektur oraz ich porównanie w kontekście analizy obrazów MRI mózgu.

5 Wyniki eksperymentów

W niniejszym rozdziale przedstawiono wyniki eksperymentów przeprowadzonych z wykorzystaniem dwóch architektur uczenia głębokiego: konwolucyjnej sieci neuronowej VGG19 oraz modelu transformerowego Swin Transformer. Ocenę skuteczności modeli przeprowadzono na wspólnym zbiorze testowym, z wykorzystaniem identycznych metryk jakości.

5.1 Metryki jakości klasyfikacji

Tabela 1 przedstawia porównanie podstawowych metryk jakości klasyfikacji uzyskanych przez oba modele. Wartości zostały obliczone na podstawie raportów klasyfikacji wygenerowanych dla zbioru testowego.

Tabela 1: Porównanie metryk jakości klasyfikacji dla badanych modeli

Model	Accuracy	Precision (macro)	Recall (macro)	F1-score (macro)
VGG19	0.85	0.84	0.84	0.83
Swin Transformer	0.99	0.99	0.99	0.99

Model Swin Transformer osiągnął bardzo wysokie wartości wszystkich analizowanych metryk, znacząco przewyższając klasyczną architekturę VGG19. Szczególnie widoczna jest różnica w wartości miary F1-score (macro), która uwzględnia równomiernie skuteczność klasyfikacji dla wszystkich klas.

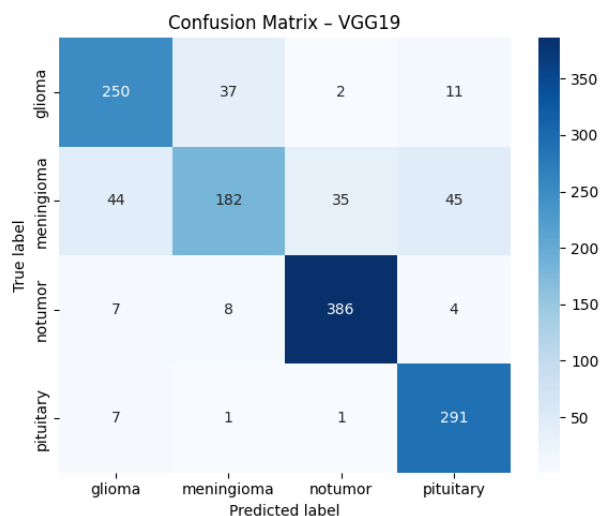
5.2 Analiza wyników dla poszczególnych klas

Szczegółowe raporty klasyfikacji wskazują, że model VGG19 wykazywał największe trudności w rozróżnianiu klasy *meningioma*, dla której uzyskano najniższą wartość czułości (recall). W przypadku modelu Swin Transformer wszystkie

klasy zostały rozpoznane z bardzo wysoką skutecznością, a wartości precyzji i czułości dla każdej z nich były zbliżone do jedności.

5.3 Macierze pomyłek

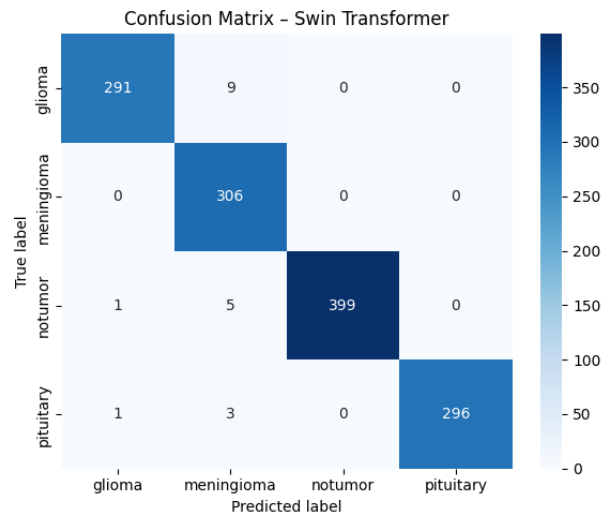
Na rysunkach 1 oraz 2 przedstawiono macierze pomyłek dla modeli VGG19 oraz Swin Transformer. Analiza macierzy pomyłek pozwala na szczegółową ocenę rodzaju popełnianych błędów klasyfikacyjnych.



Rysunek 1: Macierz pomyłek dla modelu VGG19

W przypadku modelu VGG19 widoczna jest większa liczba błędnych klasyfikacji pomiędzy klasami *glioma* oraz *meningioma*, co może świadczyć o podobieństwie cech wizualnych tych typów nowotworów.

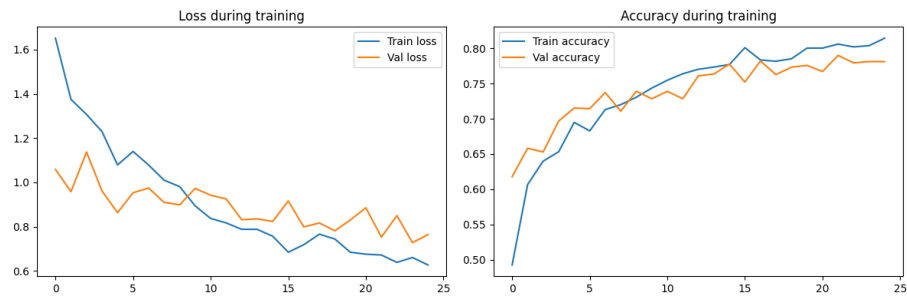
Model Swin Transformer charakteryzował się znacznie mniejszą liczbą błędnych klasyfikacji. Większość próbek została poprawnie przypisana do odpowiednich klas, a liczba pomyłek była marginalna i nie koncentrowała się wokół jednej konkretnej klasy.



Rysunek 2: Macierz pomyłek dla modelu Swin Transformer

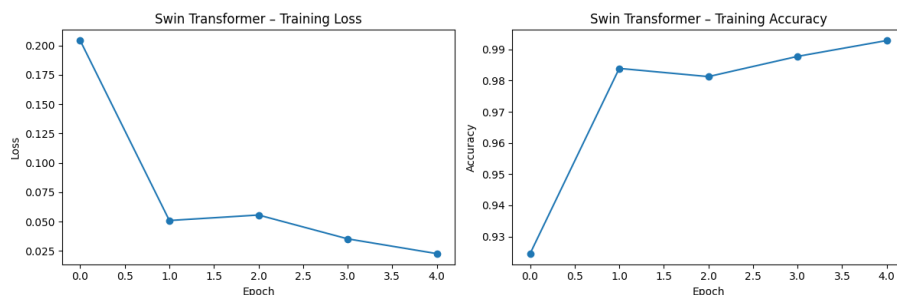
5.4 Przebieg procesu uczenia

Na rysunkach 3 oraz 4 przedstawiono przebieg procesu uczenia obu modeli w postaci zmian wartości funkcji straty oraz dokładności klasyfikacji w kolejnych epokach treningowych.



Rysunek 3: Przebieg procesu uczenia modelu VGG19

Dla modelu VGG19 zaobserwowano stopniowy spadek wartości funkcji straty oraz wzrost dokładności klasyfikacji w kolejnych epokach treningowych. Proces uczenia był stabilny, jednak tempo poprawy skuteczności malało w późniejszych etapach treningu.



Rysunek 4: Przebieg procesu uczenia modelu Swin Transformer

W przypadku modelu Swin Transformer zauważalna była bardzo szybka konwergencja procesu uczenia. Już po pierwszych epokach model osiągał wysoką dokładność klasyfikacji, a dalszy trening prowadził jedynie do niewielkich, lecz stabilnych popraw wyników.

6 Dyskusja

Uzyskane wyniki eksperymentów wskazują na wyraźne różnice pomiędzy klasyczną konwolucyjną siecią neuronową VGG19 a architekturą transformerową Swin Transformer w zadaniu klasyfikacji obrazów MRI mózgu. Model Swin Transformer osiągnął znacznie wyższe wartości wszystkich analizowanych metryk jakości, w tym dokładności klasyfikacji oraz miary F1-score w ujęciu makro, co świadczy o jego wysokiej zdolności generalizacji.

Jednym z kluczowych obserwowanych zjawisk była bardzo szybka konwergencja procesu uczenia modelu Swin Transformer. Już po kilku epokach treningowych model osiągał wysoką skuteczność klasyfikacji, podczas gdy dalsze zwiększanie liczby epok prowadziło jedynie do niewielkich, lecz stabilnych popraw wyników. W przypadku modelu VGG19 proces uczenia przebiegał wolniej i wymagał większej liczby epok w celu osiągnięcia stabilnych rezultatów.

Analiza macierzy pomyłek wykazała, że model VGG19 miał największe trudności w rozróżnianiu klas *glioma* oraz *meningioma*, co może wynikać z podobieństwa cech wizualnych tych typów nowotworów na obrazach MRI. Model Swin Transformer charakteryzował się natomiast znacznie mniejszą liczbą błędnych klasyfikacji oraz bardziej równomiernym rozkładem błędów pomiędzy klasami, co potwierdza jego zdolność do modelowania globalnych zależności w obrazie.

Należy jednak zwrócić uwagę na większą złożoność obliczeniową modelu Swin Transformer w porównaniu do sieci VGG19. Proces treningu transformera był bardziej czasochłonny oraz wymagał większych zasobów sprzętowych, co może stanowić ograniczenie w praktycznych zastosowaniach klinicznych. Z tego względu wybór architektury powinien uwzględniać kompromis pomiędzy skutecznością klasyfikacji a kosztem obliczeniowym.

Do ograniczeń przeprowadzonych eksperymentów należy zaliczyć stosunkowo

niewielki rozmiar zbioru danych oraz brak walidacji na niezależnym zbiorze zewnętrznym. Ponadto nie przeprowadzono optymalizacji hiperparametrów modeli, co w przyszłych badaniach mogłoby przyczynić się do dalszej poprawy uzyskanych wyników.

7 Wnioski

Na podstawie przeprowadzonych eksperymentów sformułowano następujące wnioski:

- Architektura transformerowa Swin Transformer osiągnęła wyraźnie wyższą skuteczność klasyfikacji obrazów MRI mózgu w porównaniu do klasycznej sieci konwolucyjnej VGG19.
- Model Swin Transformer charakteryzował się bardzo szybką konwergencją procesu uczenia, osiągając wysoką dokładność już po kilku epokach treningowych.
- Sieć VGG19, mimo prostszej architektury i mniejszych wymagań obliczeniowych, wykazywała większą liczbę błędnych klasyfikacji, szczególnie pomiędzy klasami o zbliżonych cechach wizualnych.
- Analiza macierzy pomyłek potwierdziła, że modele transformerowe lepiej radzą sobie z uchwyceniem globalnych zależności w obrazach MRI.
- W przyszłych pracach zasadnym kierunkiem badań byłoby rozszerzenie eksperymentów o większe zbiory danych, walidację zewnętrzną oraz optymalizację hiperparametrów badanych modeli.