

Analiza uogólnienia modelu segmentacji guzów mózgu na zróżnicowanych zbiorach danych MRI

Bartłomiej Wójcik

BARTEK03W@GMAIL.COM

*Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
Koszykowa 75, 00-662 Warszawa*

Kornel Tłaczała

KORNELTLACZALA@GMAIL.COM

*Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
Koszykowa 75, 00-662 Warszawa*

Tomasz Żywicki

TOMASZ.ZYWICKI03@GMAIL.COM

*Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
Koszykowa 75, 00-662 Warszawa*

Editor:

Abstract

W niniejszym projekcie przeanalizowano możliwości uogólnienia zwycięskiego modelu segmentacji guzów mózgu z zawodów BraTS 2021. Model został uruchomiony na zestawach danych innych niż oryginalny zbiór treningowy, w tym BraTS2024-Pre, BraTS2024-Post oraz BraTS-Africa. Celem analizy było (1) sprawdzenie, czy model zachowuje swoją skuteczność na nowych danych oraz (2) zbadanie, czy jakość segmentacji różni się istotnie między podgrupami pacjentów — ze względu na typ guza, wiek czy płeć. Wyniki pokazały wysoką jakość segmentacji na danych przed leczeniem, znaczący spadek skuteczności na danych po leczeniu oraz wyższe wartości wskaźnika DICE dla starszych pacjentów. Nie zaobserwowano wyraźnych różnic między płciami. Praca stanowi krok w kierunku oceny odporności modeli segmentacyjnych na różnice pomiędzy zbiorami danych i populacjami pacjentów.

1 Wprowadzenie

Segmentacja obrazów rezonansu magnetycznego (MRI) w kontekście guzów mózgu jest jednym z kluczowych problemów medycznego przetwarzania obrazów. Jej dokładność ma istotne znaczenie zarówno dla diagnozy, jak i planowania leczenia. Zawody BraTS (Brain Tumor Segmentation Challenge) od lat stanowią punkt odniesienia w ocenie jakości modeli automatycznej segmentacji. W 2021 roku zwycięski model oparty na architekturze nnU-Net osiągnął bardzo wysokie wyniki na danych konkursowych.

Jednak skuteczność modeli głębokiego uczenia często drastycznie spada w warunkach innych niż te, na których były trenowane — zjawisko to określa się jako brak generalizacji. Celem niniejszej pracy jest empiryczna analiza tej kwestii poprzez zastosowanie wspomnianego modelu na nowych, zróżnicowanych zbiorach danych.

1.1 Metryka oceny: współczynnik Dice

Do oceny jakości segmentacji wykorzystano współczynnik Dice (Dice Similarity Coefficient, DSC), który mierzy zgodność między maską predykowaną przez model a maską referencyjną (ekspercka). Współczynnik ten definiuje się jako:

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|}$$

gdzie X i Y to zbiory pikseli (lub voxelów) segmentacji modelu i maski referencyjnej odpowiednio. Wartość DSC mieści się w przedziale od 0 do 1, gdzie 1 oznacza idealne dopasowanie.

1.2 Pytania badawcze

W pracy postawiono następujące pytania badawcze:

1. Czy model segmentacyjny zachowuje skuteczność na danych innych niż zbiór treningowy?
2. Czy jakość segmentacji różni się istotnie pomiędzy podgrupami pacjentów — ze względu na typ guza (GBM vs LGG), wiek oraz płeć?

W dalszej części raportu przedstawiamy odpowiedzi na powyższe pytania, poparte wynikami eksperymentów.

2 Powiazane prace

Segmentacja struktur anatomicznych na obrazach rezonansu magnetycznego (MRI) jest jednym z kluczowych zastosowań głębokiego uczenia w medycynie. W odpowiedzi na potrzeby automatyzacji procesu konfiguracji modeli segmentacyjnych, Isensee et al. zaproponowali nnU-Net — system, który samodzielnie dostosowuje architekturę, preprocessing, augmentację oraz strategię trenowania do specyfiki danego zadania (Isensee et al., 2021a). nnU-Net nie jest nową architekturą, lecz zestandaryzowana rama, która przekształca klasyczny U-Net w konkurencyjne rozwiązanie bez konieczności ręcznego strojenia hiperparametrów. Model ten zdobył pierwsze miejsca w wielu konkursach segmentacyjnych, w tym BraTS, pokazując dużą skuteczność i ogólność.

W pracy *Extending nnU-Net for Brain Tumor Segmentation* autorzy przedstawili rozszerzenie podstawowej wersji nnU-Net z zastosowaniem do zadania segmentacji guzów mózgu w ramach konkursu BraTS 2021. Wśród najważniejszych modyfikacji znalazły się: zaawansowane techniki augmentacji danych, rozbudowana strategia ensemble obejmująca modele 2D, 3D-full resolution oraz 3D-low resolution, a także wykorzystanie test-time augmentation i selektywnego post-processingu (Isensee et al., 2022). Podejście to umożliwiło osiągnięcie najwyższych wyników w konkursie, potwierdzając skuteczność automatycznego pipeline’u wspieranego przez dodatkowe techniki poprawiające uogólnianie.

W naszym projekcie korzystaliśmy z wersji bazującej na tym rozszerzonym podejściu. W szczególności interesowało nas, jak model zachowuje się w sytuacji, gdy dane wejściowe pochodzą z rozproszonych, różnorodnych źródeł (np. dane po leczeniu lub z innych populacji). Choć oryginalna praca skupiała się na jednym zestawie danych konkursowych, nasze

badania stanowią naturalne rozszerzenie tej linii prac, testując odporność modelu na zmiany rozkładu danych i warunków obrazowania.

3 Metody

3.1 Dane

Wykorzystaliśmy cztery zbiory danych MRI z segmentacją guzów mózgu:

- BraTS 2021 – dane wykorzystywane podczas treningu oryginalnego modelu,
- BraTS 2024-Pre – dane obrazowe pacjentów przed leczeniem (pre-treatment),
- BraTS 2024-Post – dane obrazowe po leczeniu (post-treatment),
- BraTS Africa – dane z populacji afrykańskiej.

Wszystkie zbiory zawierały segmentacje trzech podregionów guza: enhancing tumor (ET), tumor core (TC) oraz peritumoral edema (ED), co pozwoliło na bezpośrednie porównanie wyników na tych samych kategoriach anatomicznych.

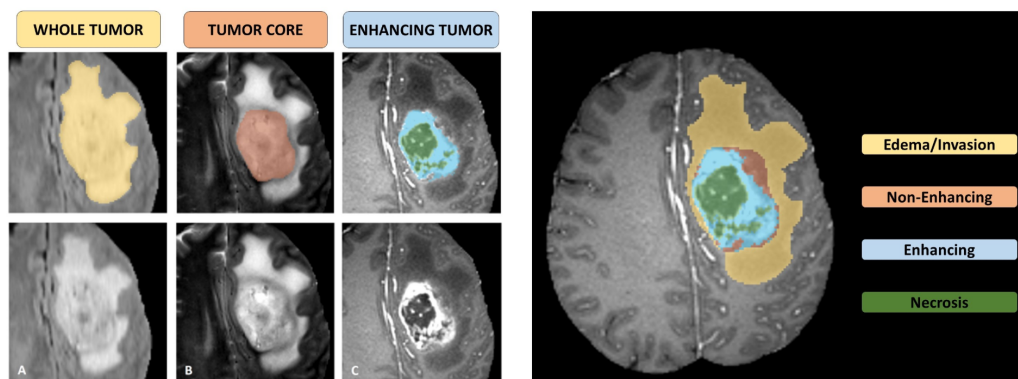


Figure 1: Wizualizacja segmentacji guza z podziałem na jego regiony.

3.2 Predykcja

Segmentacje zostały wygenerowane z wykorzystaniem wytrenowanego modelu dostępnego publicznie jako rozszerzenie nnU-Net. Nie przeprowadzono dodatkowego treningu ani fine-tuningu. Wszystkie predykcje wykonano z użyciem GPU na lokalnym środowisku testowym.

3.3 Ocena jakości

Do oceny jakości segmentacji wykorzystano metrykę Dice (DSC). Wyniki agregowano dla poszczególnych zbiorów danych i podgrup pacjentów.

4 Wyniki

Dataset	Tumor Core	Edema	Enhancing Tumor	Whole Tumor
BraTS2021	0.770	0.899	0.885	0.950
BraTS2024-Pre	0.810	0.901	0.889	0.924
BraTS2024-Post	0.135	0.681	0.357	0.618
BraTSAfrica	0.585	0.817	0.779	0.885

Table 1: Średnie wartości współczynnika Dice dla różnych regionów guza w różnych zbiorach danych

Na rysunkach 2–4 przedstawiono rozkłady wartości współczynnika Dice dla poszczególnych zbiorów danych oraz różnych regionów guza. Model uzyskuje najwyższą jakość segmentacji na oryginalnym zbiorze BraTS 2021 oraz BraTS 2024-Pre, natomiast na danych po leczeniu (BraTS 2024-Post) widać wyraźny spadek skuteczności, szczególnie dla regionu Tumor Core” i Enhancing Tumor”. Dane BraTSAfrica cechują się umiarkowanymi wynikami, co wskazuje na pewne wyzwania związane z uogólnieniem modelu na populację afrykańską. W przypadku zbioru BraTSAfrica, przyczyna umiarkowanych wyników mogą być również różnice w protokołach obrazowania (np. inny sprzęt MRI, odmienne parametry akwizycji).

Natomiast przyczyn gorszych wyników segmentacji na zbiorze BraTS 2024-Post można doszukiwać się w kilku aspektach. Najistotniejszym czynnikiem jest obecność zmian pooperacyjnych, które znacząco różnią się morfologicznie od obrazów przed leczeniem. Model trenowany wyłącznie na danych przedoperacyjnych (BraTS 2021) nie miał możliwości nauczania się reprezentacji charakterystycznych dla blizn, obszarów po resekcji, zmian pozapalnych czy atypowego wzorca wzmocnienia kontrastowego. Powoduje to trudności w precyzyjnym rozróżnianiu struktur guza, zwłaszcza w regionach Tumor Core” i Enhancing Tumor”, dla których Dice spada do odpowiednio 0.13 i 0.36.

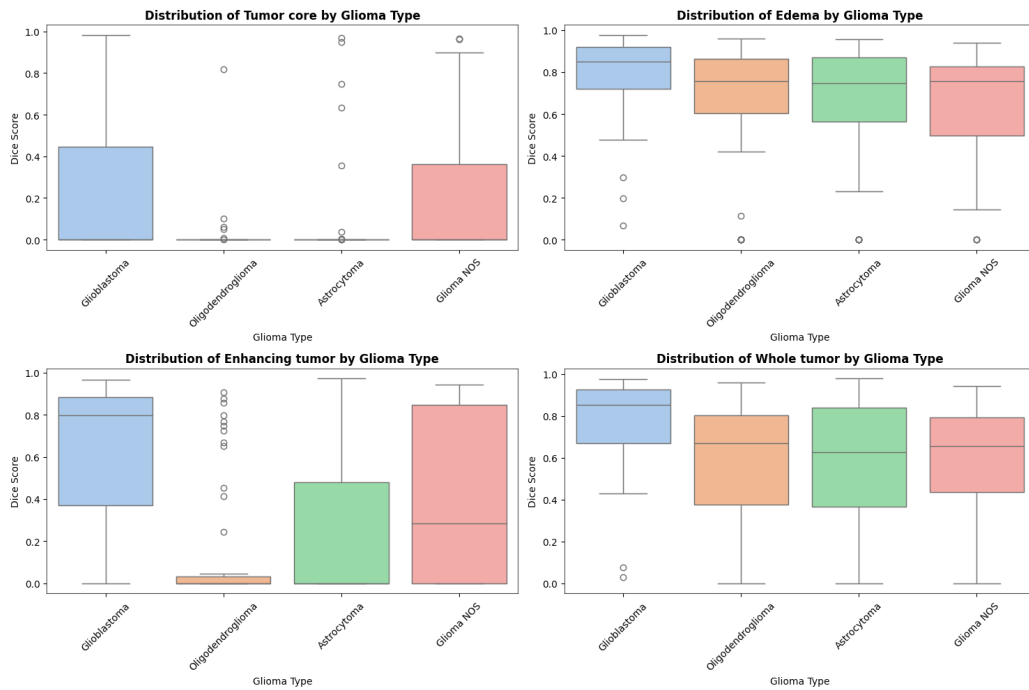


Figure 2: Rozkład wartości Dice dla różnych regionów mózgu i rodzajów nowotworów

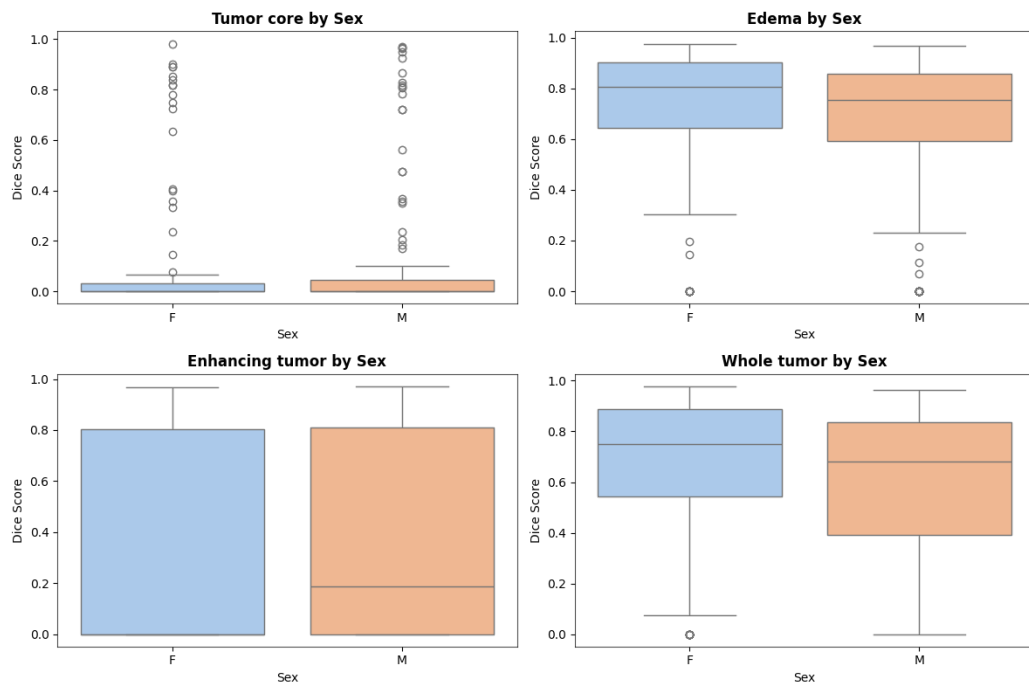


Figure 3: Rozkład wartości Dice dla różnych regionów mózgu i płci

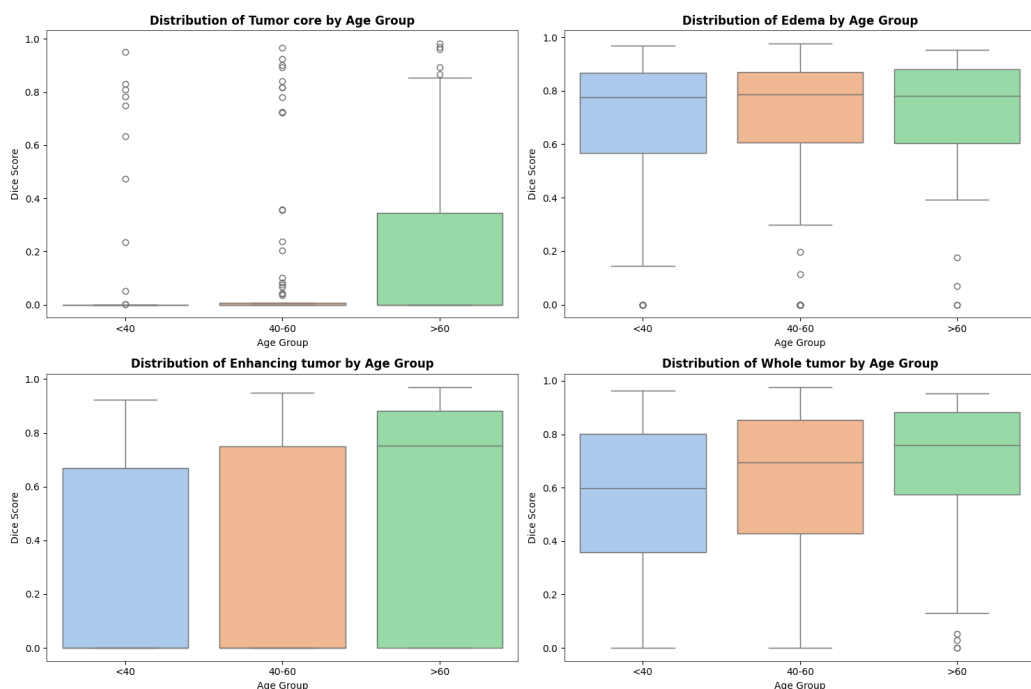


Figure 4: Rozkład wartości Dice dla różnych regionów mózgu i grup wiekowych

5 Dyskusja i kluczowe kontrybucje

Przeprowadzone eksperymenty potwierdzają ograniczenia modeli segmentacyjnych w kontekście uogólniania na nowe zbiory danych, szczególnie w sytuacji obrazów po leczeniu. Jest to ważne z punktu widzenia zastosowań klinicznych, gdzie często dostępne dane różnią się od zbiorów treningowych.

Kluczowe kontrybucje raportu to:

- Empiryczna analiza uogólnienia modelu BraTS 2021 na różne podzbiory danych, w tym nowe populacje i stany kliniczne.
- Wskazanie na spadek skuteczności segmentacji po leczeniu, co wymaga dalszych badań i adaptacji modeli.
- Wskazanie ograniczeń stosowania modeli uczonych na ustandaryzowanych zbiorach publicznych w rzeczywistych warunkach klinicznych.
- Wstępna obserwacja różnic jakości segmentacji w zależności od wieku pacjentów.

6 Wnioski

Model segmentacyjny BraTS 2021 zachowuje wysoka jakość na danych przed leczeniem oraz na oryginalnym zbiorze treningowym, jednak jego skuteczność spada na danych po leczeniu. Różnice pomiędzy płciami są nieistotne, natomiast starsi pacjenci uzyskują lepsze wyniki segmentacji.

Wyniki potwierdzają, że skuteczność modeli segmentacyjnych znacząco zależy od zgodności danych testowych z danymi treningowymi. Szczególnie widoczny jest problem w przypadku obrazów po leczeniu, gdzie obecność zmian pooperacyjnych prowadzi do znacznego spadku jakości segmentacji, zwłaszcza w regionach Tumor Core” i Enhancing Tumor”. Analogicznie, w przypadku zbioru BraTSAfrica obserwujemy obniżenie skuteczności, prawdopodobnie wynikające z różnic w protokołach obrazowania, cechach populacyjnych oraz potencjalnych różnicach w anotacjach.

Uzyskane rezultaty podkreślają znaczenie tworzenia bardziej zróżnicowanych zbiorów treningowych oraz konieczność rozwijania technik umożliwiających adaptację modeli do nowych kontekstów klinicznych i technologicznych. Bez zastosowania strategii takich jak transfer learning, fine-tuning czy adaptacja domenowa, modele segmentacyjne mogą być niewystarczające w warunkach rzeczywistych, co może ograniczać ich zastosowanie w praktyce klinicznej.

Bibliografia

References

- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18, 203–211.
- Isensee, F., Maier-Hein, K. H. (2022). Extending nnU-Net for Brain Tumor Segmentation. *arXiv preprint arXiv:2201.XXXXX*.