Dominika Wyszyńska
Mikołaj Chomanski

Etap 1. projektu badawczego, SIWY 2025L

Temat Pracy: Wykrywanie anomalii w obrazach medycznych oraz wyjaśnianie ich genezy

Wprowadzenie

Celem naszego projektu jest algorytm, który:

1. Wykrywa anomalie na na obrazach medycznych - w szczególności skany 3D.
2. Znajduje lokalizacje anomalii w przestrzeni z wykorzystaniem Saliency, GradCAM oraz LIME.
3. Optymalizuje architekturę/trening/inferencję sieci wykorzystanej do zadania z wykorzystaniem np. transfer learningu, knowledge distilation, metod atrybucji danych.

W projekcie zamierzamy skupić swoją uwagę na chorobach/urazach kolan (w tym zerwanie więzadła krzyżowego czy choroby zwyrodnieniowe). W ramach zadania planujemy wykorzystać poniższe zbiory danych:

- *https://www.kaggle.com/datasets/sohaibanwaar1203/kneemridataset* - zbiór danych, który zawiera skany 3D kolan. Każdemu zapisowi przypisano stan więzadła krzyżowego przedniego (od zdrowych do tych całkowicie zerwanych).

- *https://huggingface.co/datasets/rajpurkarlab/3DReasonKnee* - zbiór danych, który zawiera skany 3D kolan w celu oceny choroby zwyrodnieniowej stawów.

- *https://stanfordaimi.azurewebsites.net/datasets/bface6fc-7859-47d7-a1c8-022cd6b17419* - zbiór danych pochodzący ze Stanfordu, składa się ze zdjęć 3D kolan, do których przypisano diagnozy (etykiety) pozyskane ręcznie z raportów klinicznych. Zbiór obejmuje ponad 1000 badań, w tym konkretne urazy, takie jak zerwanie więzadła krzyżowego oraz uszkodzenie łąkotki.

Planowany stack technologiczny

- Python
- PyTorch
- Biblioteki: Captum, SHAP, LIME, Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn
- Narzędzia takie jak wandb lub/oraz MLflow do śledzenia eksperymentów
- black, ruff
- Środowisko uv + venv
- Streamlit / Gradio do wizualizacji wyników i uruchamiania przez użytkownika
- pytest

Proponowane algorytmy i modele

W ramach projektu konieczne jest zaimplementowanie i przetestowanie dwóch typów algorytmów: modeli klasyfikacyjnych oraz metod wyjaśnialności, które pozwolą na interpretację ich wyników.
Wedle artykułu [1] wykorzystanie całego modelu 3D, z nie tylko zdjęć skanu, pozwala osiągać lepsze wyniki w predykcji. Poniżej przedstawiono modele/algorytmy, które będą testowane w roli klasyfikatorów. Ich zadaniem będzie predykcja występowania schorzeń na podstawie danych ze skanów 3D oraz wyznaczenie obszaru objętego chorobą/urazem.

- Prosta sieć CNN z warstwami 3D konwolucji - prosta sieć wielowarstwowa z kilkoma warstwami konwolucji 3D, a następnie kilkoma warstwami perceptronów wielowarstwowych. Konwolucja 3D różni się od konwolucji 2D jednym więcej wymiarem przestrzeni. W naszej pracy posłuży jako “baseline”, do którego będziemy porównywać pozostałe modele.
- 3D Res-Net - klasyczny model rezydualny, składa się z kilku bloków rezydualnych. Każdy z tych bloków zawiera kilka warstw konwolucyjnych oraz połączenie rezydualne z wejściem tego bloku. W naszej implementacji planujemy zastąpić warstwy konwolucyjne 2D warstwami konolucyjnymi 3D (tak naprawdę dodamy jeszcze jeden wymiar).
- 3D Vision Transformer - transformer do analizy obiektów 3D można wykorzystać na dwa sposoby. W pierwszym na wejście transformera wprowadzamy elementy siatki (np. płaszczyzny lub krawędzie) i otrzymujemy predykcje. W drugim podejściu dzielimy przestrzeń na określoną ilość sześcianów i kodujemy każdy z tych sześcianów, aby kodowania mówiły, co w danym sześcianie się znajduje. Tak zakodowane sześciany wprowadzamy na wejście transformera i otrzymujemy predykcje. Kluczowe jest, aby kodowania były dyskretne, ponieważ z takimi wartościami lepiej radzi sobie model (tzn. każdy sześcian kodowany jest jako słowo np. z informacją, że w tym sześcianie znajduje się rzepka). Podejście analogiczne do podejścia zastosowanego do obrazów 2D.

Wykorzystamy również gotowe modele do klasyfikacji innych schorzeń ze skanu MRI. Wykorzystamy transfer wiedzy, aby sprawdzić jak łatwo wykorzystać gotowe modele 3D do predykcji na podobnych zagadnieniach. Jeden z modeli, które możemy wykorzystać to *https://huggingface.co/TencentMedicalNet/MedicalNet-Resnet10*.

Jednym z ulepszeń, które możemy zastosować jest użycie modelu, który wyznacza sześcian, w którym znajduje się większość kolan, co pozwoli zmniejszyć wielkość skanów, a co za tym idzie zmniejszy rozmiar modeli.

Poniżej znajdują się proponowane metody wyjaśnialności. Zostaną one zastosowane do powyższych klasyfikatorów, aby zwizualizować i zinterpretować proces podejmowania przez nie decyzji.

- GradCAM - technika wizualizacji decyzji podejmowanych przez model, wykorzystująca gradienty i mapy generowane na podstawie tych gradientów. Metoda ta propaguje gradient do ostatniej warstwy konwolucyjnej i z wyjścia tej warstwy tworzone są mapy cech. Metoda jest niezależna od architektury i dla obrazów 2d generuje heatmapę o rozdzielczości niższej niż oryginalne zdjęcie. Zakładamy że dla implementacji 3D wyjściem będzie pokolorowana “chmurka”.
- Silency - najprostsza metoda wyjaśniania predykcji modelu, wykorzystująca gradient. Wykorzystamy ją jako porównanie dla metody GradCAM. Główną różnicą w porównaniu do innych metod gradientowych jest sposób propagacji wstecz przez funkcje ReLu. W tej metodzie gradient “przepuszczamy” w tył, tylko wtedy gdy wartość aktywacji ReLu jest większa od 0. Wyjściem będzie również “chmurka”.
- LIME - metoda wyjaśniania. W procesie wyjaśniania perturbuje piksele i sprawdza, jak ważne były one do predykcji. Sama metoda nie jest przeznaczona do zdjęć i prawdopodobnie działa jeszcze gorzej dla obiektów 3D. W naszej pracy posłuży jako podstawowy algorytm wyjaśnialny, na bazie którego będziemy wyjaśniać.

Planowana funkcjonalność programu

Program będzie miał za zadanie określić, czy dany skan 3D kolana jest zdrowy bądź czy jest uszkodzony przez chorobę (np. zerwane więzadło czy choroby zwyrodnieniowe). Jeżeli skan wskazuje, że kolano nie jest zdrowe, program powinien za pomocą zaimplementowanych metod XAI w przybliżeniu określić rejon kolana objęty chorobą/urazem oraz na jakiej podstawie podjął takie decyzje.

Planowany zakres eksperymentów

- Wybór najlepszego modelu bazowego klasyfikującego skany 3D kolan;
- Wytrenowanie wybranego modelu bazowego klasyfikującego skany 3D kolan (testowanie parametrów modelu oraz wybór najlepszych);
- Adaptacja metod wyjaśniania predykcji do wejść 3D;
- Wyznaczenie wyjaśnień dla przykładów testowych i ocenienie jakości tych wyjaśnień.

Proponowany harmonogram projektu

lp.	Zakres dat	uwagi	Planowane działania
1	10.11 - 14.11		prototyp
2	15.11 - 26.11		Athensy/Kolokwia
3	27.11 - 02.12		Setup Środowiska i test prosty modeli dla obiektów 3D
4	03.12 - 09.12		Stworzenie zaawansowanego modelu do klasyfikacji i trening
5	10.12 - 16.12		Trening modelu
6	17.12 - 21.12		Modyfikacja metod wyjaśniania do modeli 3D
7	22.12 - 02.01		święta
8	03.01 - 07.01		Implementacja wizualizacji
9	07.01 - 14.01		Finalizacja Sprawozdania/Artykułu

Bibliografia

[1] Singh, S.P.; Wang, L.; Gupta, S.; Goli, H.; Padmanabhan, P.; Gulyás, B. “3D Deep Learning on Medical Images: A Review”. Sensors 2020, 20, 5097. *https://doi.org/10.3390/s20185097*

[2] Kim, Seonggyeom, and Dong-Kyu Chae. "Exmeshcnn: An explainable convolutional neural network architecture for 3d shape analysis." Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 2022. *https://doi.org/10.1145/3534678.3539463*

[3] M. R. Karim et al., "DeepKneeExplainer: Explainable Knee Osteoarthritis Diagnosis From Radiographs and Magnetic Resonance Imaging," in IEEE Access, vol. 9, pp. 39757-39780, 2021, doi: 10.1109/ACCESS.2021.3062493.

[4] Kokkotis, Christos, et al. "Leveraging explainable machine learning to identify gait biomechanical parameters associated with anterior cruciate ligament injury." Scientific Reports 12.1 (2022): 6647.

[5] Hegde, Vishakh, and Reza Zadeh. "Fusionnet: 3d object classification using multiple data representations." arXiv preprint arXiv:1607.05695 (2016).

[6] Zhou, Hong-Yu, et al. "nnformer: Volumetric medical image segmentation via a 3d transformer." *IEEE transactions on image processing* 32 (2023): 4036-4045.