Wyjaśnialny system do klasyfikacji obrazów medycznych

Michał Durkalec - 263917@student.pwr.edu.pl
Wiktoria Majewska - 263981@student.pwr.edu.pl

1 Wstęp

Celem projektu było zaproponowanie architektury systemu informatycznego, który pozwala na klasyfikację obrazów RTG klatki piersiowej z wykorzystaniem sieci neuronowych. System ma wspomagać lekarzy w diagnostyce chorób poprzez automatyczną klasyfikację obrazów, a także dostarczać informacji o tym, jakie cechy obrazu zostały wykorzystane do podjęcia decyzji.

Głównym założeniem projektu nie jest stworzenie dokładnych modeli, ale eksploracja zagadnienia Explainable AI (XAI) [5, 9] i możliwości implementacji mechanizmów XAI w aplikacjach dla użytkownika końcowego.

1.1 Motywacja

Zastosowanie głębokich sieci neuronowych w medycynie ma ogromny potencjał, ale wymaga specjalnych rozwiązań związanych z bezpieczeństwem i etyką. W szczególności problematyczny jest brak zrozumienia mechanizmów decyzyjnych głębokich sieci neuronowych co generuje dalsze problemy regulacyjne i etyczne. [4]

Efektywne *XAI* może pomóc w budowaniu zaufania i transparentości modeli dla użytkowników końcowych (pacjentów i lekarzy), a co za tym idzie, znacząco przyspieszyć procesy diagnostyczne [8]. Ponadto wykorzystanie modeli głębokich w medycynie otwiera nowe możliwości w przygotowaniu spersonalizowanych planów terapii, co może znacząco poprawić jakość życia pacjentów [10].

1.2 Zakres projektu

Ze względu na ograniczenia czasowe projektu, zakres implementacji został znacząco zawężony, a głównym celem było przeanalizowanie możliwości i potrzeb użytkowników końcowych w zakresie wyjaśnialności modeli. W szczególności w zakres projektu wchodziły następujące zagadnienia:

- Architektura systemu informatycznego
- Proof-Of-Concept (PoC) dla klasyfikacji obrazów RTG klatki piersiowej
- Wizualizacja wyników klasyfikacji w końcowej aplikacji
- Analiza wyzwań związanych z wyjaśnialnością modeli

2 Architektura systemu

System oparty jest o architekturę klient-serwer, gdzie warstwa serwerowa odpowiada za przetwarzanie obrazów i klasyfikację, a warstwa klienta za interakcję z użytkownikiem. Taki podział jest konieczny, ponieważ przetwarzanie obrazów i inferencja modeli wymaga dużej mocy obliczeniowej, a także specjalistycznych narzędzi i bibliotek.

2.1 Warstwa klienta

Prosta aplikacja webowa, która pozwala na przesłanie obrazu RTG klatki piersiowej i wyświetlenie wyniku klasyfikacji dla zalogowanych użytkowników. W dlaszej perspektywnie aplikacja może zostać rozbudowana o dodatkowe funkcjonalności lub integrację z innymi systemami informatycznymi w placówce medycznej.

Technologie: React.js, TypeScript, Material-UI

2.2 Warstwa serwera

API REST, które pozwala na przesłanie obrazu RTG klatki piersiowej, przetworzenie go i zwrócenie wyniku klasyfikacji. Serwer, w zależności od stopnia eksploatacji, może zostać zaimplementowany w architekturze mikroserwisów, co pozwoli na łatwe skalowanie i zarządzanie zasobami.

Technologie: FastAPI, Python, PyTorch

3 Proof-Of-Concept

Proof-Of-Concept (PoC) został zaimplementowany w języku Python z wykorzystaniem biblioteki PyTorch. Kod źródłowy jest dostępny jako załącznik do raportu i może być uruchomiony na dowolnej maszynie z zainstalowanym środowiskiem Python i Jupyter.

Prototyp pokazuje w jaki sposób można wykorzystać mechnizm Gradient CAM (Class Activation Mapping) [3] do wizualizacji aktywacji w sieciach konwolucyjnych. W ramach prototypu zrealizowano:

1. Przygotowanie zbioru danych
2. Dostosowanie modelu ResNet50 do klasyfikacji obrazów RTG klatki piersiowej
3. Wizualizacja aktywacji w sieci

Ze względu na ogranicznone możliwości obliczeniowe zaimplementowany model nie jest precyzyjny i nie nadaje się do zastosowań produkcyjnych; jego głównym celem jest pokazanie możliwości wizualizacji aktywacji w sieciach konwolucyjnych.

3.1 Zbiór danych

Do treningu klasyfikatora wykorzystano zbiór danych oparty na *CheXpert* [6]. Wykorzystany zbiór można pobrać z `https://www.kaggle.com/datasets/ashery/chexpert`. Zbiór zawiera 224 316 obrazów RTG klatki piersiowej z 65 240 pacjentów, z których każdy został opisany przez 14 różnych etykiet.

3.2 Model podstawowy

Do klasyfikacji obrazów RTG klatki piersiowej wykorzystano zmodyfikowany model ResNet50 [2]. ResNet jest jednym z najpopularniejszych modeli w dziedzinie klasyfikacji obrazów i jest często wykorzystywany w zastosowaniach medycznych. Ponadto ResNet jest dostępny w wielu różnych implementacjach w bibliotekach takich jak PyTorch czy TensorFlow.

3.3 Dostosowanie modelu

Model ResNet50 został dostosowany do klasyfikacji obrazów RTG klatki piersiowej poprzez zmianę warstwy wyjściowej. ResNet jest modelem przeznaczonym do klasyfikacji obrazów z bazy ImageNet [1], która zawiera obrazy z 1000 różnych klas. W związku z tym, ostatnia w pełni połączona warstwa modelu ResNet50 została zastąpiona warstwą z 14 neuronami, odpowiadającymi 14 etykietom z CheXpert.

Następnie model był trenowany na zbiorze danych CheXpert przez 1 epokę z wykorzystaniem algorytmu *AdamW* [7]. Jako funkcję straty wykorzystano *BCEWithLogitsLoss*, a jako metrykę oceny jakości klasyfikacji *AUROC* (Area Under Receiver Operating Characteristic).

Z powodu ograniczonych możliwości obliczeniowych nie podjęto próby usprawnienia modelu poprzez hiperparametryzację czy zastosowanie bardziej zaawansowanych technik regularyzacji. Dostosowany model osiągnął jakość klasyfikacji na poziomie 0.5 AUROC, co jest wynikiem losowym i nie nadaje się do zastosowań produkcyjnych.

3.4 Wizualizacja aktywacji

Do wizualizacji aktywacji w wykorzystano mechanizm Gradient CAM (Class Activation Mapping) [3]. Pierwszym krokiem działania mechanizmu jest wybór ostatniej warstwy konwolucyjnej w modelu, która zawiera informacje o lokalizacji cech w obrazie. Następnie obliczane są gradienty aktywacji w tej warstwie względem klasy docelowej, co pozwala na określenie, które obszary obrazu były najbardziej istotne dla

klasyfikacji. Obliczone gradienty są następnie agregowane w celu uzyskania mapy aktywacji, która jest nakładana na obraz wejściowy.

Do realizacji wizualizacji aktywacji wykorzystano biblioteki PyTorch, PIL.Image oraz pytorch_gradcam. W ramach prototupu możliwa jest wizualizacja aktywacji dla dowolnego obrazu RTG klatki piersiowej w obrębie wszystkich 14stu klas dostępnych w zbiorze danych CheXpert.

Ponieważ prototyp jest oparty o model o niskiej jakości, wizualizacje aktywacji nie są wiarygodne i nie nadają się do zastosowań diagnostycznych. Efektywnie, prototyp pokazuje jedynie w jaki sposób można zaimplementować mechanizm GradCAM w sieciach konwolucyjnych oraz jakiego rodzaju wyników należy oczekiwać i przetwarzać w warstwie prezentacji.

4 Wizualizacja warstwy prezentacji systemu

Tutaj mockupy

5 Analiza wyzwań związanych z wdrożeniem systemu

Wprowadzenie aplikacji wspomagającej diagnostykę chorób klatki piersiowej z wykorzystaniem Explainable AI (XAI) wiąże się z istotnymi wyzwaniami etycznymi i prawnymi. Kluczowe obszary to:

- **Ochrona danych medycznych przed nieautoryzowanym dostępem:** Dane medyczne są uznawane za szczególnie wrażliwe i wymagają najwyższego poziomu ochrony. Zgodnie z art. 9 ust. 1 RODO, przetwarzanie danych dotyczących zdrowia jest co do zasady zabronione, chyba że zachodzą określone wyjątki. W kontekście aplikacji mobilnej konieczne jest zapewnienie odpowiednich środków technicznych i organizacyjnych, aby chronić dane przed nieautoryzowanym dostępem, utratą czy modyfikacją.
- **Zgodność z RODO w przypadku przetwarzania danych osobowych:** RODO nakłada na administratorów danych obowiązek informowania pacjentów o celach i podstawach prawnych przetwarzania ich danych, a także o przysługujących im prawach. W przypadku aplikacji diagnostycznej konieczne jest spełnienie obowiązku informacyjnego wobec użytkowników oraz zapewnienie możliwości realizacji ich praw, takich jak prawo dostępu do danych czy ich sprostowania.
- **Ryzyko błędów klasyfikacji oraz wpływ takich błędów na decyzje diagnostyczne lekarzy:** Algorytmy AI mogą popełniać błędy, które w kontekście medycznym mogą prowadzić do poważnych konsekwencji dla pacjentów. Zgodnie z art. 4 ustawy o zawodach lekarza i lekarza dentysty, lekarz ma obowiązek wykonywać zawód zgodnie ze wskazaniami aktualnej wiedzy medycznej oraz z należytą starannością. Dlatego aplikacja powinna być traktowana jedynie jako narzędzie wspomagające, a ostateczna decyzja diagnostyczna powinna należeć do lekarza.
- **Transparentność działania modelu dzięki mechanizmom XAI:** Wykorzystanie mechanizmów XAI pozwala na wyjaśnienie decyzji podejmowanych przez model AI, co zwiększa zaufanie użytkowników do systemu. Transparentność działania algorytmu jest kluczowa, aby lekarze mogli zrozumieć podstawy decyzji i ocenić ich wiarygodność.
- **Ograniczenie aplikacji do narzędzia wspomagającego, a nie zastępującego lekarza w procesie diagnozy:** Zgodnie z art. 37 ustawy o zawodach lekarza i lekarza dentysty, w razie wątpliwości diagnostycznych lekarz powinien zasięgnąć opinii właściwego specjalisty lub zorganizować konsylium lekarskie. Aplikacja powinna być traktowana jako wsparcie w procesie diagnostycznym, a nie jako zastępstwo dla profesjonalnej oceny medycznej. Lekarz ponosi odpowiedzialność za ostateczną diagnozę i leczenie pacjenta.

Bibliografia

[1] Jia Deng i in. „ImageNet: A Large-Scale Hierarchical Image Database". W: (2009), s. 248–255. DOI: 10.1109/CVPR.2009.5206848. URL: https://ieeexplore.ieee.org/document/5206848 (term. wiz. 26. 11. 2024).

[2] Kaiming He i in. „Deep Residual Learning for Image Recognition". W: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, s. 770–778. DOI: 10.1109/CVPR.2016.90. URL: https://arxiv.org/abs/1512.03385 (term. wiz. 26. 11. 2024).

[3] Ramprasaath R. Selvaraju i in. „Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization". W: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2017), s. 618–626. DOI: 10.1109/ICCV.2017.74. URL: https://arxiv.org/abs/1610.02391 (term. wiz. 26. 11. 2024).

[4] Rob Challen i in. „Artificial Intelligence, Bias and Clinical Safety". W: *BMJ Quality & Safety* 28.3 (2019), s. 231–237. DOI: 10.1136/bmjqs-2018-008370. URL: https://qualitysafety.bmj.com/content/28/3/231 (term. wiz. 26. 11. 2024).

[5] David Gunning i in. „XAI—Explainable Artificial Intelligence". W: *Science Robotics* 4.37 (2019). DOI: 10.1126/scirobotics.aay7120. URL: https://www.science.org/doi/10.1126/scirobotics.aay7120 (term. wiz. 26. 11. 2024).

[6] Jeremy Irvin i in. „CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison". W: *arXiv preprint arXiv:1901.07031* (2019). URL: https://arxiv.org/abs/1901.07031 (term. wiz. 26. 11. 2024).

[7] Ilya Loshchilov i Frank Hutter. „Decoupled Weight Decay Regularization". W: *International Conference on Learning Representations (ICLR)* (2019). URL: https://arxiv.org/abs/1711.05101 (term. wiz. 26. 11. 2024).

[8] Julia Amann i in. „Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective". W: *BMC Medical Informatics and Decision Making* 20.1 (2020), s. 310. DOI: 10.1186/s12911-020-01332-6. URL: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6 (term. wiz. 26. 11. 2024).

[9] Alejandro Barredo Arrieta i in. „Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI". W: *Information Fusion* 58 (2020), s. 82–115. DOI: 10.1016/j.inffus.2019.12.012. URL: https://arxiv.org/abs/1910.10045 (term. wiz. 26. 11. 2024).

[10] Ben Allen. „The Promise of Explainable AI in Digital Health for Precision Medicine: A Systematic Review". W: *Journal of Personalized Medicine* 14.3 (2024), s. 277. DOI: 10.3390/jpm14030277. URL: https://www.mdpi.com/2075-4426/14/3/277 (term. wiz. 26. 11. 2024).