

Warsztaty Badawcze - Praca domowa 3

Mateusz Flis, Tomasz Modzelewski, Przemysław Olender

14 kwietnia 2022

1 Covid 19 radiography database

1.1 Główne informacje o artykule

Tytuł artykułu: Lightweight Neural Network for COVID-19 Detection from Chest X-ray Images Implemented on an Embedded System

Autorzy: Theodora Sanida, Argyrios Sideris, Dimitris Tsiktsiris and Minas Dasygenis

Data publikacji: 25.02.2022

1.2 Informacje o bazie danych w artykule

Baza danych zawiera 21165 zdjęć X-ray podzielonych na cztery grupy w zależności od choroby zbadanej wśród pacjenta:

- COVID-19 pozytywne przypadki - 3616
- Normalne, zdrowe płuca - 10192
- Lung Opacity - Niecovidowe infekcje płuc - 6012
- Wirusowe zapalenie płuc - 1345

Wszystkie zdjęcia są rozszerzenia .png oraz mają wymiary 299x299 pixeli. Artykuł nie wspomina o metadanych, ani o ułożeniu pacjentów podczas zdjęcia x-ray. Dodatkowo, artykuł pomija wszelkie problemy ze zdjęciami, które znaleźliśmy podczas ostatniej pracy domowej. Aparatury medyczne, łańcuszki, naszyjniki występujące na wielu zdjęciach nie są wspomniane w tej pracy naukowej. Nie znaleźliśmy informacji o źle ułożonych, przyciętych zdjęciach, zbyt przyciemnionych bądź rozjaśnionych, których niestety jest dość dużo w tej bazie. Podczas badań autorzy nie wykorzystali masek do zdjęć, które są zawarte w głównej bazie. Chociaż może to być kwestia różnicy pomiędzy wersjami, gdyż maski zostały dodane do COVID - 19 Radiography Database dość niedawno.

1.3 Wyniki modelowania przedstawione w artykule

Autorzy zaprezentowali kilka modeli predykcyjnych do klasyfikacji bazy danych. Wśród przetrenowanych modeli CNN, najlepszy wynik miał MobileNetV2 - 95,4% accuracy. Dodatkowo, następujące modele osiągnęły poniższe rezultaty tej metryki:

- Bayesian CNN - 89,82%
- CoroNet - 89,6%
- CovXNet - 90,2%
- InceptionResNetV2 - 92,18%

Na początku autorzy podzielili randomowo bazę danych na 3 zestawy. Zestaw treningowy zawierał 70% wszystkich danych, walidacyjny 20%, a testowy 10%. Ponieważ większość CNN są trenowane na rozmiarach 224x224, postanowili zmienić wymiary przy użyciu Python Imaging Library (PIL). Mimo, iż dokładność klasyfikacji MobileNetV2 wynosi 71% klasyfikacji modeli top-1 takich jak VGG16 oraz VGG19, autorzy zdecydowali się na ten model przez następujące plusy: mała liczba parametrów, mała liczba obliczeń, duża efektywność, niskie wymagania siły obliczeniowej oraz krótki czas oczekiwania. Następnie wykorzystali dodatkowe trzy warstwy do podstawowego modelu CNN. Pierwszą zastosowaną warstwą jest globalne średnie grupowanie - global average pooling. Następnie użyli Rectified Linear Unit (ReLU) activation function. Ta funkcja wygląda następująco:

$$ReLU(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \quad (1)$$

Ostatnią dodatkową warstwą jest SoftMax z 4 węzłami wyjściowymi - po jednym dla każdej klasy. Poniżej przedstawiamy funkcję aktywacyjną Softmax dla $m = 4$:

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{y=1}^m e^{x_y}} \quad (2)$$

Główny model zaprezentowany w dokumencie - MobileNetV2 - osiągnął następujące wyniki metryki precision dla poszczególnych klas:

- COVID-19 pozytywne przypadki - 0.9888
- Normalne, zdrowe płuca - 0.9527
- Lung Opacity - Niecovidowe infekcje płuc - 0.9416
- Wirusowe zapalenie płuc - 0.9923
- Średnia - 0.9688
- Średnia ważona - 0.9582

2 COVID-19 CT segmentation dataset

2.1 Główne informacje o artykule

Tytuł artykułu: Covid-19 radiology — data collection and preparation for Artificial Intelligence

link do artykułu: <https://medium.com/@hbjenssen/covid-19-radiology-data-collection-and-preparation-for-artificial-intelligence-4ecece97bb5b>

Autor: Håvard Bjørke Jenssen

Data publikacji: 28.03.2020

2.2 Informacje o bazie danych w artykule

Baza danych zawiera 110 zdjęć i jest podzielona na zbiór treningowy (100 zdjęć) oraz testowy (10 zdjęć). Zdjęcia pochodzą ze strony Włoskiego Towarzystwa Radiologii Medycznej i Interwencyjnej z której zostały pobrane w formacie .JPG. Autorzy artykułu przekonwertowali zdjęcia do formatu Nifti, uspójnili rozmiar zdjęć do wielkości 512x512 pikseli, kolory zostały zmienione na czarno-białe. Znormalizowano intensywność zdjęć w odwrotnej kolejności, pobierając wartości RGB z obrazów JPG z obszarów powietrza i tłuszczu, wykorzystano to do ujednolicenia skali Hounsfielda (taki zabieg służy unormowaniu gęstości różnych ośrodków, np. powietrza w celu osiągnięcia uniwersalnego standardu). Zdjęcia zostały zanotowane ręcznie przez radiologów używając trzech etykiet: 1 = ground class opacification, 2 = consolidations, 3 = pleural effusions.

W artykule przedstawiony jest też preprocessing zdjęć oraz zmiana formatu na Nifti, autorzy nie tworzyli żadnych modeli.

2.3 Różnice między artykułem a bazą

Wszystkie informacje z artykułu zostały zawarte w bazie danych. Jedyną nieścisłość to brak oznaczeń etykiet w plikach bazy a jedynie wspomnienie ich na stronie z materiałami do pobrania.

W artykule nie ma żadnej wzmianki o metadanych, można domyślać się, że pochodzą z tego samego źródła co zdjęcia, ale nie jest wyjaśniony powód braków w danych oraz brakuje informacji o tym, dla których pacjentów wykonano kilka zdjęć i w jakim odstępie czasu to nastąpiło.

Różnice w informacjach między artykułem z bazą dotyczą głównie metadanych, co jest drugorzędna kwestią przy rozpoznawaniu zmian na zdjęciach płuc, można więc uznać, że baza jest dobrze opisana.

3 Chest X-ray images Open-i

3.1 Główne informacje o artykule

Tytuł artykułu: Preparing a collection of radiology examinations for distribution and retrieval

link do artykułu: <https://pubmed.ncbi.nlm.nih.gov/26133894/>

Autorzy: Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, Clement J McDonald

Data publikacji: 01.07.2015

3.2 Informacje o bazie danych w artykule

Artykuł podaje źródło zdjęć i opisów jakim są dwa szpitale z Indiana Network for Patient Care, jest w napisane o tym, że baza zawiera 3996 raportach radiologicznych (ponieważ tylko tyle autorzy byli w stanie ręcznie przejrzeć) odpowiadających 8121 zdjęciom. Opisy zdjęć zostały automatycznie zanonimizowane, zdjęcia również zostały poddane automatycznej anonimizacji, jednak ze względu na niewystarczającą skuteczność wymagane były ręczne poprawki. Do bazy zostały dodawane tylko raporty zawierające sekcje z powodem badania, diagnozą oraz oceną zdjęcia. Raporty były anotowane jako nienormalne - jeśli zawierały informacje o ciężkich lub przewlekłych chorobach lub normalne w przeciwnym wypadku. Następnie, jeśli zdjęciu przypisano etykiety not normal zamieniano ją na przybliżoną nazwę dolegliwości.

Dokładnie opisana jest zawartość najważniejszych części raportu - Impression i Findings.

Do artykułu dołączone są

- przykładowy raport,
- przykładowe zdjęcie z zanonimizowanych zamazanym obszarem i zdjęcie na którym widoczna jest aparatura medyczna i zęby pacjenta,
- zrzut ekranu wyniku wyszukiwania zapalenia płuc w repozytorium ze zdjęciami.
- Tabela z 10 najpopularniejszymi etykietami chorób (zapisywanych zamiast not normal).

3.3 Różnice między artykułem a bazą

Traktujemy tę bazę jako tekstową, więc skupimy się na różnicach dotyczących raportów.

Wszystkie informacje zawarte w artykule można odnaleźć w bazie danych.

Pliki XML zawierające raporty posiadają znacznie więcej informacji niż przykładowy raport dołączony do artykułu, między innymi: informacje o licencji,

autorze, tytule, typie i dacie publikacji artykułu, link do zdjęcia, podpis zawierający pozycje w jakiej zdjęcie zostało wykonane.

Jednak te informacje są nieistotne z punktu widzenia osoby tworzącej modele uczenia maszynowego. Baza jest bardzo dobrze opisana w artykule.

3.4 Artykuły korzystające z bazy

Artykuł jest cytowany 52 razy. Wykonywane były głównie zadania klasyfikacji oraz segmentacji w większość na zdjęciach, pojawiały się też analizy i porównania zdjęć oraz raportów przy użyciu NLP. Poniżej wyniki z jednego artykułu zajmującego się klasyfikacją raportów.

3.4.1 Natural Language Processing of Radiology Text Reports: Interactive Text Classification

Artykuł jest wprowadzeniem do tematyki analizy tekstów radiologicznych przy użyciu sieci neuronowych, baza z Indiana University jest jedynym źródłem danych dla modelu. Klasyfikacja sprowadzała się do określenia, czy raport informuje o chorobie czy nie (klasy normal i abnormal), wykorzystano architekturę RNN sieci neuronowych, a wyniki zaprezentowano za pomocą confusion matrix:

- TP: 202
- TN: 18
- FP: 367
- FN: 7

4 CheXpert

4.1 Główne informacje o artykule

Link do artykułu: <https://arxiv.org/pdf/1901.07031.pdf>

Tytuł artykułu: CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison

Autorzy: Jeremy Irvin i in.

Data publikacji: 21.01.2019

4.2 Informacje o bazie danych w artykule

Baza danych zawiera 224316 zdjęć rentgenowskich wykonanych u 65240 pacjentów, a także 14 etykiet (nazw chorób lub faktu braku choroby), z wartościami: pozytywne (stwierdzenie choroby), negatywne (brak choroby) bądź niepewne. Dane były gromadzone w Stanford Hospital w okresie od października 2002 r. do lipca 2017 r.

Artykuł zawiera istotne informacje dotyczące sposobu przydzielania przykładom etykiet, w tym różnice pomiędzy sposobami znakowania przykładów ze

zbiorów: treningowego, walidacyjnego i testowego (np. do znakowania zbioru walidacyjnego zatrudniano trzech radiologów, a do testowego - aż ośmiu).

W artykule nie poświęcono natomiast dostatecznej uwagi liczbie braków danych w etykietach. W artykule wspomniane są trzy wartości (negatywna, nierozstrzygająca, pozytywna), natomiast w naszej bazie danych ponad 70% wszystkich etykiet to nie wartość niepewna (-1), lecz zwykły NaN. Jest to istotna informacja, bowiem w większości przypadków zdjęcie nie będzie opisane 14 etykietami, lecz znacznie mniejszą ich liczbą, co utrudni wytrenowanie modelu. Nie poświęcono również wystarczająco dużej uwagi kwestii rozmiarów zdjęć - jak wykazano w pracy domowej nr 2 - mamy istotny odsetek zdjęć o różnych wymiarach. Ta kwestia powinna być zostać dokładniej omówiona, gdyż różne rozmiary zdjęć także mogą utrudnić trenowanie modelu. Pominięcie tych informacji było raczej celowe, możliwe, że wynikało z faktu, iż trudno w tak krótkim artykule drobiazgowo rozwodzić się nad takimi szczegółami.

4.3 Wyniki modelowania przedstawione w artykule

Autorzy wykonali zadanie klasyfikacji, testując wiele różnych konwolucyjnych sieci neuronowych (przy użyciu optymalizatora Adam), m.in. ResNet152, Inception-v4, SE-ResNeXt101, DenseNet121 - ta ostatnia osiągnęła najwyższe wyniki.

Na potrzeby trenowania wszystkie obrazy zmniejszono do rozmiarów 320×320 .

Zestaw walidacyjny składał się z 200 zdjęć od 200 różnych pacjentów, każde było niezależnie ocenione przez trzech radiologów. W zestawie testowym zdjęć było 500 i każde było klasyfikowane na podstawie ocen łącznie ośmiu radiologów.

Końcowy model uzyskał dla wszystkich etykiet wskaźnik AUC wyższy niż 0.9, z wyjątkiem niedodmy, ang. *Atelectasis*, dla której uzyskał 0.85. Ogólny wskaźnik AUC wyniósł 0.907. Wskaźnik AUPRC był najwyższy dla *Pleural Effusion* - 0.91, a najniższy dla *Consolidation* - 0.44.

Inne artykuły również korzystały z tej bazy danych, przykładowo *Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification*, opublikowany w ICCV2021, a także *Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels*, opublikowany w Science Direct. W obu przypadkach wykonano zadanie klasyfikacji wieloklasowej i uzyskano wynik AUC równy ok. 0.93.

5 Pulmonary Chest X-Ray Abnormalities

5.1 Główne informacje o artykule

Link do artykułu: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256233/pdf/qims-04-06-475.pdf>

Tytuł artykułu: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases

Autorzy: Stefan Jaeger i in.

Data publikacji: 16.11.2014

5.2 Informacje o bazie danych w artykule

Artykuł zawiera niezwykle ważne informacje (niedostępne bezpośrednio w zbiorze danych), m.in. to, że zbiór został stworzony w celu wykrywania gruźlicy. Zawiera on 662 zdjęcia frontalne, z czego 326 to zdrowe płuca, a 336 - osób chorych na gruźlicę (klasyfikacja binarna). Pochodzą one ze szpitala w Shenzhen (Chiny), były wykonywane głównie we wrześniu 2012 r. Do każdego zdjęcia dołączony jest bardzo krótki opis, który zawiera wiek i płeć pacjenta oraz stwierdzone nieprawidłowości (jeżeli takie miały miejsce).

Artykuł nie informuje o jednym drobnym szczególe - mianowicie o fakcie, iż na każdym lub prawie każdym zdjęciu widnieje litera L lub R (pokazywaliśmy to w pracy domowej 2). Jest to bardzo istotna wiadomość, którą powinniśmy uzyskać przed trenowaniem modelu.

5.3 Wyniki modelowania przedstawione w artykule

W tym konkretnym artykule brak jest informacji o wynikach jakiegokolwiek trenowania. Jednak zbiór danych był wykorzystywany w wielu innych artykułach, np. w *Deep learning-based automatic detection of tuberculosis disease in chest X-ray images* (doi.org/10.5114/pjr.2022.113435), gdzie autorzy zaproponowali konwolucyjną sieć neuronową, która uzyskała zarówno dokładność, jak i wynik F1 w wysokości 87%. Z kolei w artykule pt. *COVID-19 image classification using deep learning: Advances, challenges and opportunities*, opublikowanym w ScienceDirect, gdzie wykorzystano zdjęcia m.in. z Shenzhen do klasyfikacji pacjentów z COVID-19 (zdjęcia z Shenzhen, z uwagi na dawną datę wykonania, posłużyły jako przykłady braku tej ostatniej choroby).