

Praca Domowa 3

Bartosz Siński
Jakub Grunas
Maciej Pawlikowski

April 2022

IQ-OTH/NCCD [4]

Article:

1. Basic Information

- Data was gathered during three months of the fall 2019.
- Data was gathered in the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases
- Data set contains 1190 images representing CT scan slices of 110 cases. They are grouped into three groups: 40 cases are diagnosed with malignant cancer, 15 cases diagnosed with benign cancer, and 55 cases classified as normal cases.

2. Important information in the article not featured in the data set

- The CT scans were originally collected in DICOM format then exported to png.
- Slice thickness of 1 mm.
- Cases vary in gender, age, educational attainment, area of residence and living status.
- Patients come from different places in the middle region of Iraq, particularly, the provinces of Baghdad, Wasit, Diyala, Salahuddin, and Babylon.

3. Information about data set not contained in the article

- There was no explanation why images were converted from DICOM to png and therefore all DICOM attributes and metadata was lost.
- There was no information if chosen patients have any other diseases.
- There was no precise information about patients age or gender. Only fact that they vary was mentioned. Maybe such an information was not accessible to the authors of the article.

- No explanation why one group contains only 15 cases while other more than twice as big.

4. Tasks conducted on this dataset

In the main article that contains information about data set, detection of the lung cancer was performed. The authors used SVM to perform classification of the cases into three groups: malignant, benign and normal. During the preparation of the data authors also did segmentation to extract nodules from the images. Their classification task resulted in F score = 97.842%.

ieee8023/covid-chestxray-dataset [3]

1. Basic Information

- Data set contains 679 frontal chest X-ray images from 412 people from 26 countries.
- Data was largely compiled from public databases on websites such as Radiopaedia.org, the Italian Society of Medical, Interventional Radiology and the Hannover Medical School and other. Together there are 76 data sources !!!
- There are 326 PA, 192 AP, 161 AP supine and 82 Lateral images.

2. Important information in the article not featured in the data set

- Information about how data was gathered/scraped.
- Why the images are so different from each other. Know we know it is due to multiple data sources.

3. Information about data set not contained in the article

None

4. Tasks conducted on this data set

Article about data set proposes different tasks that can be performed on the data set. Proposed example for classification task is differentiation of COVID-19 from other causal agents such as bacteria or other viruses [7]. Other would be predicting if patient should be at the ICU (Intensive Care Unit) [2]. Although first one resulted in low scores (AUC around 0.55) and second one very well (AUC around 0.81). Attempts in predicting survival resulted in low scores (AUC around 0.55) and no method worked well.

VinDr-CXT[6]

Creation year: 2021

Data Source: Hospital 108 (H108) and the Hanoi Medical University Hospital(HMUH) in Vietnam

Data volume: 18 000

Important information not contained in data:

- all pictures are PA,
- test data set was labeled by consensus of 5 radiologists, while training set only by 3(without consensus),
- all outliers were excluded from data set,
- all patients are adults.

Information about data set not contained in the article: None

Tasks conducted on this data set:

- detection and segmentation
(A Structure-Aware Relation Network for Thoracic Diseases Detection and Segmentation , accuracy 58.1%[10]),
- high accuracy heat mapping
(Improved image classification explainability with high-accuracy heatmaps[1]),
- visual explanation of class activation map
(M-CAM: Visual Explanation of Challenging Conditioned Data set with Bias-reducing Memory[9]),
- Computer Aided Diagnosis System
(A clinical validation of VinDr-CXR, an AI system for detecting abnormal chest radiographs [5]).

UCSD-AI4H/COVID-CT[8]

Creation year: 2020

Data Source: the MedPix database, LUNA dataset, radiopaedia website, PubMed Central (PMC)

Data volume: 216 Covid positive patients(349 pictures), number of Covid negative patients not determined(397 pictures).

Important information not contained in data:

- how much data was imported from certain data sources.

Information about data set not contained in the article:

- no explanation why so many columns in data set are practically empty,
- no mention of mostly empty columns in metadata.

Tasks conducted on this data set:

- Classification (COVID-CT-Dataset: A CT Image Data set about COVID-19, F1 0.90, AUC 0.98, accuracy 0.89[8])

PadChest

1. Basic information
 - Creation years: 2009-2017
 - Data sources: San Juan Hospital
 - Amount of data: 160 868 images, 69 882 patients
 - 174 unique radiological discoveries
 - 27% manually annotated, the rest - automated
2. Important information in the article not featured in the dataset
 - Image preprocessing and conversion (from DICOM to jpg) description
 - Differences between projections description
 - Labels explained
 - Automation of the annotation process explained
3. Information not in the article

The article is very comprehensive and contains all necessary information (in my opinion)
4. Which other articles used this dataset?

This page shows 36 articles referencing Padchest dataset. Most of them describe multi-label classification tasks.

JSRT

1. Basic information
 - Creation year: 1998
 - Data sources: Japanese Radiological Society
 - Amount of data: 247 images (154 images with diseases, 93 normal)
 - Additional data in the dataset: age, sex, diagnosis
2. Important information in the article not featured in the dataset
 - Images features
 - Categorisation thresholds for changes' size feature
 - Performance measure of 20 radiologists who were given the task to detect changes and annotate the images
3. Information not in the article
 - Exact sources of the images
4. Which other articles used this dataset?
This page shows 2 articles referencing this dataset

References

- [1] Konpat Preechakul Sira Sriswasdi Boonserm Kijirikul Ekapol Chuangsuwanich. Improved image classification explainability with high-accuracy heatmaps. *IScience*, 25, 03 2022.
- [2] Joseph Paul Cohen, Lan Dao, Paul Morrison, Karsten Roth, Yoshua Bengio, Beiyi Shen, Almas Abbasi, Mahsa Hoshmand-Kochi, Marzyeh Ghassemi, Haifang Li, and Tim Q Duong. Predicting covid-19 pneumonia severity on chest x-ray with deep learning. 2020.
- [3] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong, and Marzyeh Ghassemi. Covid-19 image data collection: Prospective predictions are the future, 2020.
- [4] Hamdalla Kareem, Muayed Al-Huseiny, Furat Mohsen, Enam Khalil, Zainab Hassan, Hamdalla Fadil, and Kareem Al-Yasriy. Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset. *Indonesian Journal of Electrical Engineering and Computer Science*, 21, 03 2021.
- [5] Ngoc Huy Nguyen, Ha Quy Nguyen, Nghia Trung Nguyen, Thang Viet Nguyen, Hieu Huy Pham, and Tuan Ngoc-Minh Nguyen. A clinical validation of vindr-cxr, an ai system for detecting abnormal chest radiographs. *arXiv*, 2021.
- [6] Pham H. H. tuan linh l. Dao M. khanh Nguyen, H. Q. Vindr-cxr: An open dataset of chest x-rays with radiologist annotations (version 1.0.0). *PhysioNet*, 06 2021.
- [7] Linda Wang and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. 2020.
- [8] Xingyi Yang, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, and Pengtao Xie. Covid-ct-dataset: A ct scan dataset about covid-19. *arXiv*, 2020.
- [9] Jie Lian; Jingyu Liu; Shu Zhang; Kai Gao; Xiaoqing Liu; Dingwen Zhang; Yizhou Yu. M-cam: Visual explanation of challenging conditioned dataset with bias-reducing memory. *British Machine Vision Association (BMVA)*, 40, 11 2021.
- [10] Jie Lian; Jingyu Liu; Shu Zhang; Kai Gao; Xiaoqing Liu; Dingwen Zhang; Yizhou Yu. A structure-aware relation network for thoracic diseases detection and segmentation. *IEEE Transactions on Medical Imaging*, 40, 04 2021.