COVID-19: A Survey On Public Medical Imaging Data Resources

Roman Kalkreuth
Department of Compuer Science
TU Dortmund University
roman.kalkreuth@tu-dortmund.de
http://www.cs.tu-dortmund.de

Paul Kaufmann
Chair for Computational Intelligence
Johannes Gutenberg University Mainz
paul.kaufmann@uni-mainz.de
https://ci.bwl.uni-mainz.de/

Abstract

This regularly updated survey provides an overview of public resources that offer medical images and metadata of COVID-19 cases. The purpose of this survey is to simplify the access to open COVID-19 image data resources for all scientists currently working on the coronavirus crisis.

1   Introduction

Recent study results have show that medical imaging can be used for the diagnosis of COVID-19 [17]. Moreover, two recent studies indicate that COVID-19 is possibly better diagnosed with radiological imaging when compared to the PCR-Test [12, 16]. To classify computer tomography (CT) images of COVID-19 cases automatically, a Convolutional Neural Network has been used in [18]. The study aimed to answer the questions, if artificial intelligence technology can be used to early screen COVID-19 patients from their CT images and what is the corresponding diagnostic accuracy. The study also focused on the development of an early screening model to distinguish COVID-19 pneumonia from Influenza-A viral pneumonia and healthy cases. The experiments showed that the overall accuracy was 86.7% from the perspective of CT cases as a whole.

Recently, a call has been launched for scientists working in the field of artificial intelligence to focus their research and knowledge on the current coronavirus crisis [1]. At the moment, there is a lack of sufficiently large and publicly available records of image data. Data is, however, essential for machine learning techniques. With the intention to simplify the access to currently publicly available image data and metadata we started this survey.

2   Motivation

At present, medical capacities are severely limited in many countries. Also, the flow of information and exchange between physicians and scientists is limited due to the lockdowns in various countries. More massive data sets with image data of COVID-19 cases are, therefore, either not available or not publicly accessible. However, since these data sets are urgently needed to support the diagnosis and treatment of COVID-19 with machine learning or artificial intelligence technologies, we would like to contribute with this initiative to the full exploration and further development of the currently available potential of different technologies.

3   About the Initiative

Since the authors are currently working on the detection of COVID-19 with machine learning themselves, this survey will be updated regularly. We would be very grateful if members of the community could share sources with us, which are not yet listed here. Please report all errors to our email addresses. We will fix them as soon as possible. With this initiative, we also would like to encourage all scientists to make their data sets publicly available to the community, provided that data protection allows it. This initiative can also be supported by making it better known with a citation.

Resource/Dataset	Types of Data					Organization/Platform
	CT	X-Ray	MRT	Metadata	Case review	
COVID-19 Image Data Collection [15, 14]	Y	Y	N	Y	N	Institute for Reproducible Research
COVID-19 Open Research Dataset (CORD-19) [7]	Y	Y	U	Y	Y	Allen Institute for AI and partners
COVID-Net Open Source Initiative [8]	N	Y	N	N	N	Vision and Image Processing Lab
SIRM COVID-19 Database [11]	Y	Y	U	Y	Y	Italian Society of Radiology
COVID-19 BSTI Imaging Database [3]	Y	Y	N	Y	Y	British Institute of Radiology
COVID-19 Chest X-Ray Database [5]	N	Y	N	Y	N	Kaggle
COVID-19 Chest X-Ray [4]	N	Y	N	Y	N	Kaggle
COVID-19 CT Segmentation Dataset [6]	Y	N	N	N	N	medicalsegmentation.com
COVID-CT[19, 20]	Y	N	N	Y	N	University of California San Diego
Augmented COVID-19 X-ray Images Dataset [13]	N	Y	N	N	N	Dep. of Biomedical Systems & Informatics Engineering
coronacases.org [2]	Y	N	N	Y	Y	Radiology Artificial Intelligence
eurorad.org [9]	Y	Y	U	Y	Y	European Society of Radiology
radiopaedia.org [10]	Y	Y	U	Y	Y	Radiopaedia

Table 1: List of COVID-19 Medical Imaging Resources and Datasets

4 List of Resources

Table 1 shows a list of resources and datasets which provide publicly accessible medical imaging data of COVID-19 cases. The types of data are classified by CT, X-Ray, magnetic resonance tomography (MRT), metadata of the corresponding patient/case, and case review. The respective classes are annotated with **Y** for *yes*, **N** for *no* and **U** for *unknown*.

5 Note of the Authors

We would like to point out not to claim any form of diagnostic performance of models until clinical studies have been performed.

References

[1] Call to Action to the Tech Community on New Machine Readable COVID-19 Dataset. https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/. Accessed: 2020-04-07.

[2] Coronacases. https://coronacases.org. Accessed: 2020-04-07.

[3] COVID-19 BSTI Imaging Database. https://bsticovid19.cimar.co.uk/. Accessed: 2020-04-06.

[4] COVID-19 Chest X-Ray. https://www.kaggle.com/bachrr/covid-chest-xray. Accessed: 2020-04-07.

[5] COVID-19 Chest X-Ray Database. https://www.kaggle.com/tawsifurrahman/covid19-radiography-database. Accessed: 2020-04-07.

[6] COVID-19 CT Segmentation Dataset. http://www.medicalsegmentation.com/covid19/. Accessed: 2020-04-07.

[7] COVID-19 Open Research Dataset (CORD-19). https://pages.semanticscholar.org/coronavirus-research. Accessed: 2020-04-07.

[8] COVID-Net Open Source Initiative. https://github.com/lindawangg/COVID-Net/. Accessed: 2020-04-06.

[9] Eurorad. https://www.eurorad.org/. Accessed: 2020-04-07.

[10] Radiopedia. https://www.radiopaedia.org. Accessed: 2020-04-02.

[11] SIRM COVID-19 Database. https://www.sirm.org/category/senza-categoria/covid-19/. Accessed: 2020-04-02.

[12] AI, T., YANG, Z., HOU, H., ZHAN, C., CHEN, C., LV, W., TAO, Q., SUN, Z., AND XIA, L. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A [Report of 1014 Cases. *Radiology* (2020), 200642.

[13] ALQUDAH, A. M., AND QAZAN, S. Augmented COVID-19 X-Ray Images Dataset. https://data.mendeley.com/datasets/2fxz4px6d8/4. Accessed: 2020-04-07.

[14] COHEN, J. P., MORRISON, P., AND DAO, L. COVID-19 Image Data Collection. https://github.com/ieee8023/covid-chestxray-dataset. Accessed: 2020-04-07.

[15] COHEN, J. P., MORRISON, P., AND DAO, L. COVID-19 Image Data Collection. *arXiv 2003.11597* (2020).

[16] FANG, Y., ZHANG, H., XIE, J., LIN, M., YING, L., PANG, P., AND JI, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* (2020), 200432.

[17] LI, Y., AND XIA, L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. *American Journal of Roentgenology*, 1–7.

[18] XU, X., JIANG, X., MA, C., DU, P., LI, X., LV, S., YU, L., CHEN, Y., SU, J., LANG, G., LI, Y., ZHAO, H., XU, K., RUAN, L., AND WU, W. Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, 2020.

[19] ZHAO, J., ZHANG, Y., HE, X., AND XIE, P. COVID-CT. https://github.com/UCSD-AI4H/COVID-CT. Accessed: 2020-04-07.

[20] ZHAO, J., ZHANG, Y., HE, X., AND XIE, P. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. *arXiv preprint arXiv:2003.13865* (2020).