Reference

[1] Wang Enci, Fan Song, WU Xue-bin, Pu Xian-juan, JIAO Zheng, Nie Yongyou. Analysis of pollution emission characteristics of new energy vehicle based on GREET model [J]. Journal of Shanghai university (natural science edition), 2017,23 (05) : 810-820.

[2] Huo Hong,Zhang Qiang, Streets David G,He Kebin. Environmental implication of electric vehicles in China.[J]. Environmental science & technology,2010,44(13).

[3] Fang Jingrui. Analysis, Research and Evaluation of energy and Environmental Benefits of New Energy Vehicles [D]. Jilin University, 2009

УДК 611.08

MEDICAL BIG DATA IN BELARUS

Kapitonov A.A.[1], Bahretsova M.V.[1], Tsynkevich U.V.[1], Trushel N.A.[2], Kachur S.L.[3]
[1]Belarusian State Medical University, Department of Pediatrics
[2]Belarusian State Medical University, Department of Normal Anatomy
[3]Belarusian State Medical University, Department of Radiology
e-mail: [1]Kapitonov@bsmu.by, [2]normanat@bsmu.by, [3]raydiag@bsmu.by

The COVID-19 pandemic has caused millions of infection cases, which led to great economical and social losses all over the world. Due to high false-negative rate of the Reverse Transcription Polymerase Chain Reaction (RT-PCR) tests [1] and impossibility to use antibody tests for early-stage diagnostics [2], diagnosing based on radiological methods, such as Chest X-Ray (CXR) and Chest Computer Tomography (CCT), became wide-spread [3]. Researchers from different countries claim that CXR may have the same with CCT sensitivity rate, because of the same radiological signs of the COVID-19-associated pneumonia [4, 5]. This makes CXR an essential method: cheap, wide-spread, fast. It is estimated, that increase of images to analyse will be compensated by new AI-driven tools helping to classify CXR's [6]. As the area of computer vision rapidly develops aiming to assist radiologists with managing their workflow, it becomes evident that we don't have enough collected and properly prepared data for the Machine Learning process. Reported results (Table 1) don't cover the full variety of the most frequent possible pathological conditions (Table 2) [7].

Table 1 - Reported classifications

Classification	Number of articles, reporting usage of this classification
Covid pneumonia/ other pneumonia	3
Covid pneumonia/ other pneumonia/ normal image	8
Covid pneumonia/ normal image	7
Covid pneumonia/ bacterial pneumonia/ other viral pneumonia/ normal image	1
Pneumonia/ normal image	1
Normal image/ tumor/ pleural effusion/ infectious process/ other	1
Covid pneumonia/ other	2
Covid pneumonia/ other pneumonia/ normal image/ other	1

Table 2 - Real distribution of image types at 5th City Clinical Hospital (Minsk, 29.04.2020)

Type of image	Covid pneumonia	Venous stasis	Hydrothorax	Emphysem	Tumor	Normal image	Total
Amount	24	8	2	15	1	48	98
In %	24%	8%	2%	15%	1%	49%	100%

As the biggest available CXR datasets [8, 9] do not contain images with signs of COVID-19-associated pneumonia, it becomes important to collect a new topical dataset. Our research group collected 10683 CXR's from 4th, 5th & 6th City Clinical Hospitals in 3 classes:

1) Normal image

2) Pneumonia: covid and other

3) Other pathological condition: venous stasis, hydrothorax, emphysem and other

During our research were created 2 neural networks to solve the problem of CXR's classification. The model created in collaboration with the ITMO University (Saint-Petersburg, Russia) is circumstantially discussed in the article [10]. The model created in collaboration with the Laboratory of Mathematical Methods of Image Processing of the MSU University (Moscow, Russia) reached 85% accuracy. Our web-service, based on the mentioned models (http://3.21.250.7:8501/), reached great results during it's practical usage at the Urban Children's Infectious Clinical Hospital (Minsk). We report 716 right predictions from 785 attempts (91% accuracy), given by at least one neural network.

To summarize, we can declare huge potential of the medical Big Data industry in Belarus. As it was shown during our research, it is possible to collect and prepare data just in time to solve the actual problems with usage of ML&Computer vision technologies.

References

1. Böger, Beatriz, et al. "Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19." American journal of infection control (2020).

2. Deeks, Jonathan J., et al. "Antibody tests for identification of current and past infection with SARS-CoV-2." Cochrane Database of Systematic Reviews 6 (2020).

3. Wong, Ho Yuen Frank, et al. "Frequency and distribution of chest radiographic findings in COVID-19 positive patients." Radiology (2020): 201160.

4. Zhang, Jinjun, et al. "Familial cluster of COVID-19 infection from an asymptomatic." Critical Care 24.1 (2020): 1-3.

5. Schiaffino, Simone, et al. "Diagnostic performance of chest x-ray for COVID-19 pneumonia during the SARS-CoV-2 pandemic in Lombardy, Italy." Journal of thoracic imaging 35.4 (2020): W105-W106.

6. Santosh, K. C. "AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data." Journal of Medical Systems 44.5 (2020): 1-5.

7. Chen, Delong, et al. "A review of automated diagnosis of covid-19 based on scanning images." arXiv preprint arXiv:2006.05245.

8. Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.

9. Johnson, Alistair EW, et al. "MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs." arXiv preprint arXiv:1901.07042 (2019).

10. Gusarova, N. F., Lobantsev, A. A., Vatian, A. S., Kapitonov, A. A., & Shalyto, A. A. (2020). Comparative assessment of text-image fusion models for medical diagnostics. Information and Control Systems, (5), 7