Editorial

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.0869
Korean J Radiol 2024;25(3):217-219

Korean Journal of Radiology
KJR

Performing a Research Study Using Open-Source Deep Learning Models

Hyungjin Kim

Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

Keywords: Artificial intelligence; Deep learning; Prediction model; Open source; External validation

Publications on deep learning (DL)-based models for medical images are rapidly increasing. Some scientific journals now mandate that authors upload computer codes and/or models (e.g., weights) to publicly accessible repositories, such as GitHub, Bitbucket, or SourceForge. According to a blog post from the Public Library of Science [1], publicly available codes enhance understanding, support reproducibility and reuse, and increase efficiency across the entire scientific ecosystem. The development of prediction models, particularly DL models using medical images, requires considerable resources (e.g., data, labor, and time). Specifically, data collection involves image retrieval from a picture archiving and communication system, cleansing erroneous images and noisy labels, and sometimes manual or automatic lesion annotations, which are often labor-intensive procedures. Therefore, reusing the developed models for validation or clinical deployment could be an efficient research strategy compared to the repetitive development of multiple similar models. In this editorial, I share research examples that validate open-source DL models.

Received: September 6, 2023 **Revised:** October 19, 2023
Accepted: November 4, 2023
Corresponding author: Hyungjin Kim, MD, PhD, Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea
• E-mail: khj.snuh@gmail.com

External Validation of an Open-Source DL Model for CRs

Chest radiographs (CRs) are among the most widely used imaging examinations globally, and their extensive availability facilitates the early application of DL algorithms in this domain. Common DL models include segmentation and detection algorithms for lung nodules, masses, consolidations, pneumothorax etc. [2-4]. However, CRs may contain prognostic information beyond the traditional diagnostic findings, and DL models can effectively quantify this prognostic signature. For instance, Lu et al. [5] recently developed a convolutional neural network capable of predicting the long-term incidence of lung cancer for up to 12 years using publicly available CRs from a large randomized controlled trial. Their objective was to identify high-risk smokers for lung cancer CT screening. The model exhibited superior discrimination performance compared with that of the Centers for Medicare and Medicaid eligibility criteria in independent test sets.

Along with my colleagues, I conducted an external validation study of the model developed by Lu et al. [5] considering that the selection criteria for lung cancer CT screening are important in optimizing nationwide CT screening programs in Korea. The model was downloaded from the Github repository (https://github.com/vineet1992/CXR-LC), and image preprocessing was performed for CRs in accordance with the authors' instructions. In a retrospective analysis of 19488 individuals undergoing health checkups in Korea, the model showed good discrimination performance, and we demonstrated its added value to the 2021 United States Preventive Services Task Force recommendations (i.e., an update of the Centers for Medicare and Medicaid eligibility criteria) [6]. The model proved to be useful in reducing the number of screening candidates while maintaining the inclusion rate and positive predictive value for incident lung cancer [6]. Although the model was originally developed as a potential replacement for the Centers for Medicare and Medicaid eligibility criteria, we intentionally extended its application to test its added value against the updated criteria. It is worth noting that a prediction model can not only be validated precisely according to the target population and scheme of the original model development study, but can also be tested in an intentionally different target population or clinical workflow [7].

A User-Friendly Example Using Google Colab

A user-friendly example of an open-source DL model is available. Weiss et al. [8] published a CR-based DL model to predict lung-related mortality (https://github.com/AIM-Harvard/CXR-Lung-Risk). As in the previous example, this model was developed and validated using publicly available datasets from randomized controlled trials. The authors offered three ways to run their model: 1) a cloud-based approach, 2) a Dockerized version, 3) and a local setup. The cloud-based approach provided the code for the environment setup, data preparation (including preprocessing with sample images), and model inference using Google Colab (https://github.com/AIM-Harvard/CXR-Lung-Risk/blob/main/notebooks/cxr_lung_risk_mwe.ipynb), a cloud-based Jupyter notebook environment. Users with minimal coding proficiency can follow the codes provided to obtain model outputs using their CRs. External validation studies can be performed on various populations, including individuals undergoing cancer screening, patients with chronic pulmonary diseases, and those who have recovered from COVID-19 pneumonia. As mentioned previously, prognostic signatures extracted using DL models can be tested in intentionally different target populations.

Segmentation Models

Numerous open-source DL models are available for use in various imaging modalities and tasks. For example, DL algorithms have greatly advanced body segmentation [9-12], enabling the accurate quantification of organ dimensions. This allows the analysis of the area or volume of organs, fat, and muscle as imaging markers for diagnosing diseases and predicting patient outcomes. Segmented organ images can be used as inputs for separate DL models. The open-source models are https://github.com/QIMP-Team/MOOSE and https://github.com/wasserth/TotalSegmentator.

Technical Tips for External Validation of Open-Source DL Models

A caveat is the need to check whether the DL model inputs or outputs require transformation. Failure to apply the necessary transformations can lead to incorrect model outputs and reduced performance. Similarly, it is essential to verify and adhere to the image preprocessing steps, including windowing, cropping, and resizing.

External validation studies employing open-source DL models usually do not require cutting-edge graphics processing unit-equipped workstations. The DL models for CRs introduced in this editorial can be executed on Google Colab. However, if the number of images exceeds a few thousand, it is advisable to opt for a local setup after downloading the model weights.

Key Steps for Utilizing Open-Source DL Models from Public Repositories

1) When reviewing research articles related to prediction models, assess whether the models and/or their weights have been shared on public repositories.
2) Clone open-source models to your local repository or directory.
3) Create your dataset, which includes images for external validation.
4) For external validation, one may choose to strictly adhere to the target population and indications specified in the original study or intentionally explore broader applications for the model.
5) Follow the data preprocessing steps, which may involve variable transformations, employed in the original model development study.
6) Assess whether model calibration or recalibration is required.

CONCLUSION

External validation and/or an extended application study of open-source DL models is prudent and highly efficient. It not only reduces time and cost investment but also leverages the collective knowledge and expertise of a broader community.

Korean Journal of Radiology
KJR

Conflicts of Interest

H.K. received consulting fees from RadiSen; holds stock and stock option in Medical IP.

ORCID ID

Hyungjin Kim
https://orcid.org/0000-0003-0722-0033

Funding Statement

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00207978). However, the funders had no role in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the article for publication.

Acknowledgments

English editing of this article was done by the chatGPT (GPT-3.5; OpenAI, San Francisco, CA, USA).

REFERENCES

1. Morton L. Uphold the code: how complete, detailed and open code can enhance understanding, improve reproducibility, and change the shape of the research article [accessed on September 6, 2023]. Available at: https://theplosblog.plos.org/2022/05/uphold-the-code/
2. Hwang EJ, Goo JM, Nam JG, Park CM, Hong KJ, Kim KH. Conventional versus artificial intelligence-assisted interpretation of chest radiographs in patients with acute respiratory symptoms in emergency department: a pragmatic randomized clinical trial. *Korean J Radiol* 2023;24:259-270
3. Huh JE, Lee JH, Hwang EJ, Park CM. Effects of expert-determined reference standards in evaluating the diagnostic performance of a deep learning model: a malignant lung nodule detection task on chest radiographs. *Korean J Radiol* 2023;24:155-165
4. Yoo H, Kim EY, Kim H, Choi YR, Kim MY, Hwang SH, et al. Artificial intelligence-based identification of normal chest radiographs: a simulation study in a multicenter health screening cohort. *Korean J Radiol* 2022;23:1009-1018
5. Lu MT, Raghu VK, Mayrhofer T, Aerts HJWL, Hoffmann U. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann Intern Med* 2020;173:704-713
6. Lee JH, Lee D, Lu MT, Raghu VK, Park CM, Goo JM, et al. Deep learning to optimize candidate selection for lung cancer CT screening: advancing the 2021 USPSTF recommendations. *Radiology* 2022;305:209-218
7. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-W73
8. Weiss J, Raghu VK, Bontempi D, Christiani DC, Mak RH, Lu MT, et al. Deep learning to estimate lung disease mortality from chest radiographs. *Nat Commun* 2023;14:2797
9. Lee JH, Choi SH, Jung KJ, Goo JM, Yoon SH. High visceral fat attenuation and long-term mortality in a health check-up population. *J Cachexia Sarcopenia Muscle* 2023;14:1495-1507
10. Lee YS, Hong N, Witanto JN, Choi YR, Park J, Decazes P, et al. Deep neural network for automatic volumetric segmentation of whole-body CT images for body composition assessment. *Clin Nutr* 2021;40:5038-5046
11. Lee MH, Zea R, Garrett JW, Graffy PM, Summers RM, Pickhardt PJ. Abdominal CT body composition thresholds using automated AI tools for predicting 10-year adverse outcomes. *Radiology* 2023;306:e220574
12. Pickhardt PJ, Summers RM, Garrett JW. Automated CT-based body composition analysis: a golden opportunity. *Korean J Radiol* 2021;22:1934-1937