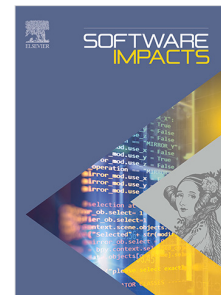


Journal Pre-proof

COV-ADXS: An Automated Detection System using X-ray images, deep learning, and XGBoost for COVID-19

Sharif Hasani, Hamid Nasiri



PII: S2665-9638(21)00093-2
DOI: <https://doi.org/10.1016/j.simpa.2021.100210>
Reference: SIMPA 100210

To appear in: *Software Impacts*

Received date : 14 December 2021
Revised date : 20 December 2021
Accepted date : 22 December 2021

Please cite this article as: S. Hasani and H. Nasiri, COV-ADXS: An Automated Detection System using X-ray images, deep learning, and XGBoost for COVID-19, *Software Impacts* (2021), doi: <https://doi.org/10.1016/j.simpa.2021.100210>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

COV-ADSX: An Automated Detection System using X-ray Images, Deep Learning, and XGBoost for COVID-19

Sharif Hasani¹, Hamid Nasiri^{2,*}

¹Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

²Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran

Correspondence should be addressed to Hamid Nasiri; h.nasiri@aut.ac.ir

Abstract

Following the COVID-19 pandemic, scientists have been looking for different ways to diagnose COVID-19, and these efforts have led to a variety of solutions. One of the common methods of detecting infected people is chest radiography. In this paper, an Automated Detection System using X-ray images (COV-ADSX) is proposed, which employs a deep neural network and XGBoost to detect COVID-19. COV-ADSX was implemented using the Django web framework, which allows the user to upload an X-ray image and view the results of the COVID-19 detection and image's heatmap, which helps the expert to evaluate the chest area more accurately.

Keywords

Chest X-ray Images; COVID-19; Deep Neural Networks; DenseNet169; XGBoost

Code metadata

Nr	Code metadata description	
C1	Current code version	V2.1.4
C2	Permanent link to code/repository used for this code version	https://github.com/sharifhasani/COV_ADSX
C3	Permanent link to reproducible capsule	-
C4	Legal code license	GNU General Public License v3.0
C5	Code versioning system used	git
C6	Software code languages used	Python
C7	Compilation requirements, operating environments and dependencies	Python 3.6 or later
C8	If available, link to developer documentation/manual	-
C9	Support email for questions	h.nasiri@aut.ac.ir

Software metadata

Current software version	2.1.4
Permanent link to executables of this version	https://github.com/sharifhasani/COV_ADSX
Legal Software License	GNU General Public License v3.0
Operating System	Microsoft Windows 7 (or later) Mac OS 10.12.6 (Sierra or later) Linux
Installation requirements & dependencies	4 GB of memory 1 GB of free disk space

Introduction

COVID-19 virus was first reported in Wuhan, China, in late 2019 and spread rapidly throughout the world [1]–[4]. The symptoms of people with COVID-19 included fever, cough, sore throat, headache, fatigue, muscle ache, and shortness of breath [5], [6]. As mentioned, one of the most common approaches for detecting COVID-19 in individuals is through PCR test; however, diagnosing people with this method is time-consuming, and the results have a high level of false-negative errors [7], [8]. Alternatively, chest radiographic imaging methods, such as computer tomography and chest X-ray (CXR) imaging, can play a vital and effective role in the timely diagnosis and treatment of this disease [7], [9].

According to the previous studies [10]–[12], X-ray images of patients infected with COVID-19 have important and useful information for detecting this virus. So, these images were employed in the software introduced in this paper (i.e., COV-ADSX). COV-ADSX uses the algorithm proposed by Nasiri and Hasani [13], through which image features were extracted using DenseNet169 [14] Deep Neural Network (DNN), and the extracted features were given as input

to the XGBoost algorithm to perform the classification. In the method proposed by Nasiri and Hasani, a pre-trained DNN was employed, and the network was not trained. COV-ADSX receives an X-ray image of a person's chest and uses a DNN to extract its features. It then gives the extracted features to the trained XGBoost to determine if the person was infected with COVID-19 or not.

XGBoost

Extreme Gradient Boosting (XGBoost) is a gradient boosting algorithm enhanced for efficiency, versatility, and scalability [15]–[17]. In recent years, XGBoost has been widely used by researchers, and it has shown impressive performance in a variety of Machine Learning (ML) challenges [18], [19]. XGBoost was proposed by Chen and Guestrin [20] as an ensemble algorithm based on gradient boosted decision trees [21]. Using a normalized objective function helps XGBoost reduce the complexity of the model and prevent overfitting [22]–[24].

Software features

COV-ADSX is based on deep learning and the XGBoost algorithm. In this software, deep learning is applied in the context of web applications. Due to the power of Python language in deep learning, this language and Django web framework [25] were employed to implement COV-ADSX. The implemented software uses Python version 3.8.8 and Django framework version 3.2.8.

COV-ADSX included four main steps as follows: 1) receiving the user's image and sending it to the deep learning model (Fig. 1); 2) extracting the features of the image using DenseNet169; 3) giving the extracted features as input to the XGBoost algorithm and performing the classification, i.e., detecting whether the person has infected with COVID-19 or not; 4) using the Gradient-based Class Activation Mapping (Grad-CAM) algorithm [26], specifying the decision area on the heatmap, and displaying it to the user (Fig. 2).

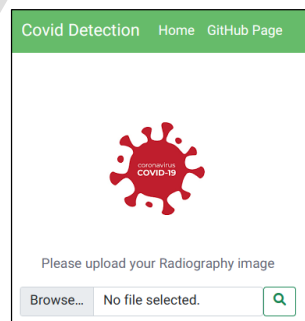


Fig. 1. Home page of COV-ADSX.

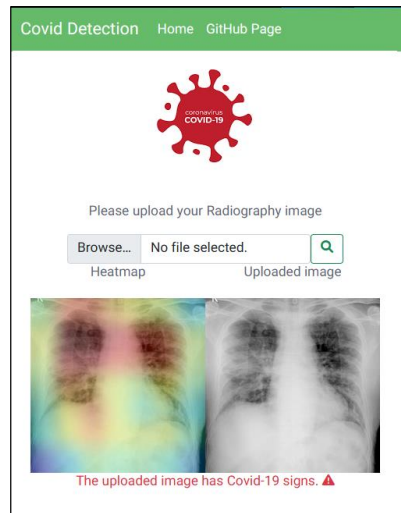


Fig. 2. The result of COV-ADSX.

In the first step, the user uploads his/her radiographic image. The image received from the user was then passed to the deep learning model to extract its features. Note that the deep learning model's weights were saved as a file and loaded into the software. The extracted features were then passed to the XGBoost algorithm to perform the detection. The XGBoost algorithm was already trained, and its trained model was employed in COV-ADSX. This software has a high speed and displays the result to the user within a maximum of 10 seconds. According to the results obtained in [13], the accuracy of the model used in COV-ADSX based on the ChestX-ray8 [27] dataset was 98.23%. After the image was detected by the XGBoost algorithm, the Grad-CAM algorithm was utilized to show the decision area on a heatmap in the last step. Heatmaps are very important to radiologists since helping them to examine the chest area more accurately.

Impact overview

As mentioned in the previous sections, COV-ADSX can diagnose people infected with the COVID-19 virus using radiographic images. COV-ADSX has a very simple user interface and can be used by all classes of society. Anyone, who has his/her CXR image, can use this software to check whether he/she is infected by COVID-19 or not. As can be seen in Fig. 1, to use the software, the user only needs to upload his/her radiographic image, and the results are then displayed to him/her (Fig. 2).

COV-ADSX provides useful information to medical specialists, especially radiologists. Its first advantage is diagnosing COVID-19 infection in a person, which normally needs much time with PCR test, in a very short time with high accuracy. In addition, it displays a heatmap to the expert

to show which parts of the input radiographic image the deep learning model is more focused on and can thus serve as an initial screening tool.

COV-ADSX can be connected to third-party applications in hospitals and medical centers. So, when a patient goes to the hospital and a CXR image is taken, the image automatically is given as input to the COV-ADSX, and the software determines if the person was infected with COVID-19 or not. Then the COVID-19 test result is recorded in the patient file. Moreover, the specified decision area on the heatmap is also stored in the patient file and is shown to the medical specialist. **To further evaluate the performance of COV-ADSX, it was tested on a limited number of patients in Shafa Hospital of Semnan (Iran). The experimental results indicated that COV-ADSX achieves reliable results in the real environment.**

Since COV-ADSX uses XGBoost for classification and XGBoost implements parallel processing [15], COV-ADSX has a lower computational cost than other ML methods. Furthermore, XGBoost has better performance than other ML techniques like Random Forest (RF) [28] and Support Vector Machine (SVR) [29], which leads to the high accuracy of COV-ADSX. The limitations and drawbacks of the COV-ADSX include using a limited number of COVID-19 X-ray images (i.e., 125 samples) for the training of the DNN model, which may decrease the generalization of software. Moreover, since XGBoost has too many hyperparameters that need to be tuned, its tuning process is challenging, and overfitting is possible if parameters are not tuned properly. Other ML methods like RF and SVR have fewer hyperparameters and are easier to tune. Another limitation of the COV-ADSX is that it only accepts CXR images as input and does not support other radiography images like chest CT. This feature can be added to the future version of the software.

Conclusion

In this paper, COV-ADSX is introduced. COV-ADSX can diagnose people infected with COVID-19 using radiographic images of a person's chest, which is impossible by the non-specialist. On the other hand, this type of imaging as the most accessible type can be easily done in most medical centers. The deep learning model used in COV-ADSX based on the Chest-Xray6 dataset has an accuracy of 98.23%. Since COV-ADSX does not require DNN training, it has a high execution speed. Displaying the heatmap of the decision area to the user is another benefit of this software that can help radiologists better diagnose COVID-19.

References:

- [1] F. Wu *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020, doi: 10.1038/s41586-020-2008-3.
- [2] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, 2020.
- [3] N. Jha *et al.*, "Deep Learning Approach for Discovery of In Silico Drugs for Combating COVID-19," *J. Healthc. Eng.*, vol. 2021, 2021.
- [4] M. Smith and F. Alvarez, "A machine learning research template for binary classification problems and shapley values integration," *Softw. Impacts*, vol. 8, p. 100074, 2021.
- [5] T. Singhal, "A review of coronavirus disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. 4, pp. 281–286, 2020.
- [6] L. Cirrincione *et al.*, "COVID-19 pandemic: Prevention and protection measures to be adopted at the workplace," *Sustainability*, vol. 12, no. 9, p. 3603, 2020.
- [7] Z. Y. Zu *et al.*, "Coronavirus disease 2019 (COVID-19): a perspective from China," *Radiology*, vol. 296, no. 2, pp. E15–E25, 2020.
- [8] P. Huang *et al.*, "Use of chest CT in combination with negative RT-PCR assay for the 2019 novel coronavirus but high clinical suspicion," *Radiology*, vol. 295, no. 1, pp. 22–23, 2020.
- [9] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, S. Mirjalili, and M. K. Khan, "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms," in *Multimodal Image Exploitation and Learning 2021*, 2021, vol. 11734, p. 117340E.
- [10] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, 2020.
- [11] J. F.-W. Chan *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [12] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.
- [13] H. Nasiri and S. Hasani, "Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost," *arXiv Prepr. arXiv:2109.02428*, 2021, [Online]. Available: <http://arxiv.org/abs/2109.02428>.
- [14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [15] H. Nasiri, A. Homafar, and S. Chehreh Chelgani, "Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using an explainable artificial intelligence," *Results Geophys. Sci.*, vol. 8, p. 100034, 2021, doi: <https://doi.org/10.1016/j.ringps.2021.100034>.
- [16] H. Dong, D. He, and F. Wang, "SMOTE-XGBoost using Tree Parzen Estimator optimization for copper flotation method classification," *Powder Technol.*, vol. 375, pp. 174–181, 2020.
- [17] Z. Zhang, Y. Huang, R. Qin, W. Ren, and G. Wen, "XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: Experiment study and modelling," *J. Manuf. Process.*, vol. 64, pp. 30–44, 2021.
- [18] H. Nasiri and S. A. Alavi, "A novel framework based on deep learning and ANOVA feature selection method for diagnosis of COVID-19 cases from chest X-ray Images," *arXiv Prepr. arXiv:2110.06340*, 2021, [Online]. Available: <http://arxiv.org/abs/2110.06340>.
- [19] M. Rapp, "BOOMER—An algorithm for learning gradient boosted multi-label classification rules," *Softw. Impacts*, vol. 10, p. 100137, 2021.
- [20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*

- Mining*, 2016, pp. 785–794.
- [21] S. C. Chelgani, H. Nasiri, and M. Alidokht, “Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A ‘conscious-lab’ development,” *Int. J. Min. Sci. Technol.*, 2021.
 - [22] S. Chehreh Chelgani, H. Nasiri, and A. Tohry, “Modeling of particle sizes for industrial HPGR products by a unique explainable AI tool- A ‘Conscious Lab’ development,” *Adv. Powder Technol.*, vol. 32, no. 11, pp. 4141–4148, 2021.
 - [23] K. Song, F. Yan, T. Ding, L. Gao, and S. Lu, “A steel property optimization model based on the XGBoost algorithm and improved PSO,” *Comput. Mater. Sci.*, vol. 174, p. 109472, 2020.
 - [24] W. Zeng, Q. Wang, Z. Xia, Z. Li, and H. Qu, “Application of XGBoost Algorithm in The Detection of SARS-CoV-2 Using Raman Spectroscopy,” in *Journal of Physics: Conference Series*, 2021, vol. 1775, no. 1, p. 12007.
 - [25] J. Forcier, P. Bissex, and W. J. Chun, *Python web development with Django*. Addison-Wesley Professional, 2008.
 - [26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
 - [27] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.
 - [28] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
 - [29] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, V. Vapnik, and others, “Support vector regression machines,” *Adv. Neural Inf. Process. Syst.*, vol. 9, pp. 155–161, 1997.

- COV-ADSX automatically detects COVID-19 from chest X-ray images.
- COV-ADSX employs a deep neural network (i.e., DenseNet169) and XGBoost to detect COVID-19.
- COV-ADSX uses gradient-based class activation mapping (Grad-CAM) to represent the decision area on a heatmap.

Declaration of interests

- ☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- ☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: