SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

Sustainable Machine Intelligence Journal

SCIENCES FORCE
International Academic Publisher

Comments on "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images"

Sami Lababidi[1,*]

[1] School of Engineering and Applied Science, University of Colorado, Boulder, CO 80309, USA; Sami.lababidi@colorado.edu

Event	Date
Received	01-06-2023
Revised	29-01-2024
Accepted	10-03-2024
Published	30-03-2024

Abstract: In an early study by (Pan et al., 2020), the authors proposed a deep learning framework; called COVIDX-Net to support radiologists in the process of automatic diagnosis of COVID-19 infection from X-ray images. In this comment, we argue that Pan et al. 's paper include three failings that may impact the performance of models available in COVIDX-Net leading to erroneous results and incorrect conclusions. First, the study lacks a clear and distinct framework, instead employing a conventional approach of training and testing a set of pre-existing deep learning models. Second, the study lies in the utilization of a very small dataset for training complex deep learning models, with no information regarding the source and annotation process. Finally, the training deep learning models is inconsistent and suffer from overfitting. The study under inspection represents a troubling example of (un/intended) exploitation of the urgency surrounding the COVID-19 pandemic for self-serving purposes, particularly to accrue citations without obeying the ethical publishing practices or demonstrating due diligence.

1. Introduction

The rapid spread of COVID-19 has placed an unprecedented burden on healthcare systems worldwide, necessitating the development of innovative diagnostic approaches. Deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have shown promise in the automated detection of COVID-19 from medical imaging data. COVIDX-Net, as proposed by Pan et al. [1], is one such framework designed to diagnose COVID-19 from X-ray images with high accuracy. While investigating the methodology and results of a COVIDX-Net, this commentary aims to highlight potential shortcomings and limitations inherent in COVIDX-Net in terms of dataset biases, model generalizability, and validation strategies. This collectively aim to give the scholars in depth understanding of misleading conclusions and challenges that might limit applying COVIDX-Net as diagnostic tools in clinical practice [1-2].

2. Comments on Data

The statement that "the dataset used in the study was provided by Dr. Joseph Cohen and Dr. Adrian Rosebrock" raises valid concerns regarding the transparency and validity of the dataset. Without comprehensive information about the source of the dataset, including details about the corresponding institution, annotators, their affiliations, and expertise in the field of radiology or medical imaging, the reliability and credibility of the data are called into question. The lack of transparency regarding the dataset's origins and the expertise of those involved in its curation raises concerns about potential biases, inconsistencies, or inaccuracies in the annotations and labels. Besides, without information about the institutional review board (IRB) approval or ethical considerations involved in the dataset collection process, it is difficult to assess whether the data collection procedures adhere to ethical standards and patient privacy regulations [2].

Given the critical role of high-quality and well-annotated datasets in training accurate and reliable DL models, it is imperative for researchers to provide detailed documentation regarding the dataset's provenance, annotation procedures, and quality assurance measures. Without such transparency, the validity and reproducibility of the study's findings may be compromised, underscoring the importance of rigorous data governance and documentation practices in the field of medical imaging research.

The authors claimed, "The X-ray images for confirmed COVID-19 disease show a pattern of ground-glass opacification with occasional consolidation in the patchy, peripheral, and bilateral areas [10]." However, upon closer examination of this reference [3], it becomes evident that the study primarily focused on computed tomography (CT) imaging for the diagnosis of COVID-19, rather than X-ray imaging. This discrepancy between the authors' claim and the referenced study is concerning, as it raises questions about the accuracy and validity of the clinical information provided in the paper. While ground-glass opacification and consolidation are indeed characteristic features of COVID-19 pneumonia as observed in CT scans, extrapolating these findings directly to X-ray images without proper validation or evidence may lead to misinterpretations and erroneous clinical decisions. Thus, it is essential for researchers to ensure the accuracy and integrity of clinical information provided in scientific publications by referencing studies that directly investigate the imaging modality in question.

3. Comments on Methods

The absence of a detailed methodology corresponding to the purported "COVIDX-Net" framework raises significant concerns regarding the clarity and coherence of the study's experimental approach. Instead of presenting a distinct framework or methodology tailored specifically for COVID-19 diagnosis from X-ray images, the method section merely outlines a conventional pipeline involving data preparation, model training, and testing using a set of pre-existing DL models, namely VGG19, DenseNet201, InceptionV3, ResNetV2, InceptionResNetV2, Xception, and MobileNetV2. This lack of specificity regarding the development and implementation of the so-called "COVIDX-Net" framework undermines the credibility and interpretability of the study's findings. Without a clear delineation of the unique features or innovations purportedly introduced by the COVIDX-Net framework, it is challenging for readers to discern the novelty or added value of the proposed approach compared to existing methodologies. Besides, the use of pre-trained DL models without customization or optimization for the task of COVID-19 diagnosis from X-ray images raises questions about the suitability and efficacy of the chosen models for this specific application. The heterogeneity of the models employed, each with its own architectural nuances and training requirements, further complicates the interpretation of results and limits

On the other hand, in the experimental setup, the DL models are trained on a dataset consisting of only 50 X-ray images, is indeed a notable departure from the standard principles of training DL models. In DL, particularly with complex architectures like those mentioned, the general rule of thumb is that larger models require proportionally larger datasets to effectively learn meaningful patterns and avoid overfitting. However, in the case of COVIDX-Net, the disparity between the size of the model and the size of the dataset raises concerns about the potential for model overfitting and the generalizability of the results. The scenario where the size of the model is comparable to or even larger than the size of the dataset can lead to several issues. Firstly, the models may simply memorize the training data rather than learning meaningful features that generalize well to unseen data, resulting in poor performance on new X-ray images. Additionally, the risk of overfitting is heightened, as the models may learn to capture noise or idiosyncrasies specific to the training dataset rather than true characteristics of COVID-19 versus normal X-ray images. Moreover, the lack of diversity and representativeness in such a small dataset may limit the model's ability to discern subtle differences and variations in X-ray images, potentially leading to biased or unreliable predictions.

4. Comments on Results

The learning curves depicted in Figure 3 exhibit significant fluctuations, indicating inconsistency during the learning process and a lack of clear convergence. This observation raises concerns about the stability and reliability of the DL models trained in the study (overfitting). While fluctuations in learning curves are not uncommon, particularly in the early stages of model training, the severity and persistence of these fluctuations in Figure 3 suggest potential issues with model convergence and optimization. It is noteworthy that the authors did not address or discuss these limitations in the paper, nor did they provide justification for the observed fluctuations in the learning curves. This omission is concerning, as it overlooks a critical aspect of model training and evaluation, which could impact the interpretation and validity of the study's findings. Understanding the factors contributing to the erratic behavior of the learning curves is essential for diagnosing underlying issues in the training process, such as inappropriate learning rates, inadequate regularization, or insufficient data preprocessing. Ignoring such basic analysis is not acceptable at all in DL study, especially during the urgent case of COVID-19.

The assertion by the authors that "Image data augmentation was not used in this study" raises concerns, particularly given the observed limitations in terms of limited data and inconsistent training. Data augmentation is a fundamental technique employed in machine learning, especially when dealing with small datasets, to enhance model generalization and mitigate overfitting [4]. Hence, the decision not to utilize data augmentation in the study, despite the evident challenges posed by the limited dataset and inconsistent training, seems counterintuitive and may have contributed to the observed issues of overfitting and instability in model performance. Ignoring elementary methods like data augmentation, which are widely recognized for their efficacy in addressing data scarcity and enhancing model generalization, raises questions about the rigor and comprehensiveness of the experimental methodology employed in the study. It is essential for researchers to carefully consider and justify the choice of methodologies and techniques used in their studies, particularly in the context of machine learning where methodological decisions can significantly impact the validity and reliability of the results.

The concerns raised regarding the overall conduct and integrity of the study are indeed significant and merit careful consideration. While the urgency of addressing the COVID-19 pandemic has understandably prompted a surge in research efforts across various disciplines, it is imperative for researchers to uphold ethical standards and scientific rigor in their publications to ensure the validity and reliability of the findings. The exploitation of the urgency surrounding the COVID-19 pandemic for the purpose of gaining citations without due diligence and responsibility to ethical publishing practices is troubling [5]. Such practices not only undermine the integrity of scientific research but also have the potential to mislead the scientific community and impede progress in combating the pandemic. Ethical considerations, transparency, and adherence to scientific standards are paramount in the pursuit of knowledge and the dissemination of research findings, particularly in critical areas such as public health emergencies. Researchers have a responsibility to conduct their studies with integrity, rigorously evaluate their methodologies, and critically assess the implications of their findings. Furthermore, the dissemination of research findings should be accompanied by a commitment to open and transparent communication, allowing for constructive critique, replication, and validation by the scientific community [6].

5. Conclusions

This commentary deliberates the serious comments concerning the paper "COVIDX-Net: A Framework of DL Classifiers to Diagnose COVID-19 in X-Ray Images." In spite of the insistence of addressing the COVID-19 pandemic, it is concludable that the paper suffer from many failures including the non-transparency of datasets, and the rationality of the methods. Together with the limited transparency and accountability in reporting, it is notable that there is a noteworthy uncertainty about the legitimacy and reliability of the study's findings. This collectively point to case of exploit the urgency of the COVID-19 outbreak to quickly publish work (even preprints) without conforming to public ethical and methodological standards, which destabilizes the dependability of scientific research misleading the scientific community.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

References

[1]. Hemdan, E. E. D., Shouman, M. A., & Karar, M. E. (2020). Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055.
[2]. Grady, Christine. "Institutional review boards: Purpose and challenges." Chest 148, no. 5 (2015): 1148-1155.
[3]. Ng, M. Y., Lee, E. Y., Yang, J., Yang, F., Li, X., Wang, H., ... & Kuo, M. D. (2020). Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiology: Cardiothoracic Imaging, 2(1), e200034.
[4]. Moreno-Barea, F. J., Jerez, J. M., & Franco, L. (2020). Improving classification accuracy using data augmentation on small data sets. Expert Systems with Applications, 161, 113696.
[5]. Seedall, C., & Tambornino, L. (2024). Research ethics and integrity in the DACH region during the COVID-19 pandemic: balancing risks and benefits under pressure. Research Ethics, 17470161241229207.
[6]. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., ... & Schönlieb, C. B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), 199-217.