Kalpa Publications in Computing

Volume 14, 2023, Pages 174–177

Proceedings of V XoveTIC Conference. XoveTIC 2022

Analysis of Imbalanced Datasets in the Performance of Deep Learning Approaches for COVID-19 Screening from Chest X-ray Imaging: Impact of Sex and Age Factors*

Lorena Álvarez-Rodríguez[1,2], Joaquim de Moura[1,2], Lucía Ramos[1,2], Jorge Novo[1,2] and Marcos Ortega[1,2]

[1] Centro de investigación CITIC, Universidade da Coruña, A Coruña, Galicia, Spain.
[2] Grupo VARPA, INIBIC, Universidade da Coruña, A Coruña, Galicia, Spain.
(lorena.alvarezr, joaquim.demoura, l.ramos, jnovo, mortega)@udc.es

Abstract

In this work, we analysed 11 imbalance scenarios with female and male COVID-19 patients present in different proportions for the sex analysis, and 6 scenarios where only one specific age range was used for training for the age factor. In each study, 3 different approaches for automatic COVID-19 screening were used: (I) Normal vs COVID-19, (II) Pneumonia vs COVID-19 and (III) Non-COVID-19 vs COVID-19.

The present study was validated using two representative public chest X-ray datasets, allowing a reliable analysis to support the clinical decision-making process.

The results for the sex-related analysis indicate this factor slightly affects the COVID-19 deep learning-based systems, although the identified differences are not relevant enough to considerably worsen the system. Regarding the age-related analysis, this factor was observed to be influencing the system in a more consistent way than the sex factor, as it was present in all considered scenarios.

* This research was funded by: Instituto de Salud Carlos III - DTS18/00136; Ministerio de Ciencia e Innovación y Universidades, Gov. of Spain - RTI2018-095894-B-I00; Ministerio de Ciencia e Innovación, Gov. of Spain - PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva - ED431C 2020/24; Axencia Galega de Innovación (GAIN), Xunta de Galicia - N845D 2020/38; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

1 Introduction

The World Health Organization (WHO) classified the COVID-19 outbreak as a pandemic in March 2020. In the wake of the global COVID-19 pandemic, which resulted in more than 575 million confirmed cases and 6.3 million deaths (WHO Coronavirus (COVID-19) Dashboard), the need for quick, accurate, and secure means of detecting and tracking respiratory infections has never been more apparent. Since these diseases have a relatively high rate of transmission due to their unique characteristics, as they are easily spread through the air, early diagnosis and evaluation of the progression of these patients is crucial because many of them, in their most severe stages, can cause symptoms including acute respiratory failure, necessitating the use of assisted breathing devices or admission to an intensive care unit (ICU).

One of the most popular methods for examining the potentially impacted areas is to use various lung imaging modalities, such as chest X-rays, in order to investigate lung involvement in greater depth. To identify and categorize the many pathological structures visible on the chest X-ray image, a thorough analysis is required, which should be carried out by a qualified expert with plenty of expertise. In this regard, it is crucial to have a set of computational approaches that enable in-depth analysis of a chest X-ray image for diagnostic reasons, especially in the current pandemic scenario.

Deep learning techniques are today without a doubt valuable tools for medical imaging analysis. However, in order to employ the produced systems in an actual environment, these methods need a lot of data. This issue, known as data scarcity, affects even more widely studied and prevalent diseases like cancer or pneumonia, where public datasets are few and, in some cases, unbalanced, containing only specific patient types. This problem was commented by Cirillo *et al*. (Cirillo, et al., 2020) in their work, as they describe how biased systems produce discriminatory results in the medical field. They focus on the sex and gender factors, as they consider these aspects to affect diseases, risks, treatments, symptoms, etc. In the work of Larrazabal *et al*. (Larrazabal, Nieto, Peterson, Milone, & Ferrante, 2020) , the authors analysed how imbalance related to gender slightly biases deep learning systems when diagnosing some lung pathologies and abnormalities through chest X-ray images, even though observed worsening was not large. In the work of Vidal *et al*. (Vidal, de Moura, Novo, & Ortega, 2021) the authors proposed a methodology that attempts to alleviate this data scarcity problem in the COVID-19 domain by a two-step knowledge transfer to obtain a robust system able to segment lung regions from portable X-ray devices despite the scarcity of samples and lesser quality. However, regardless of all the developments, the volume of articles and research, the urgency, and the lack of COVID-19 chest X-ray images, to our knowledge, no such study, specifically for sex and age, has yet been carried out for COVID-19.

2 Materials and methods

In this study, we performed a comprehensive analysis of sex and age factors in the COVID-19 datasets (Álvarez-Rodríguez, de Moura, Novo, & Ortega, 2022). For this purpose, we analyzed 3 different computational approaches for COVID-19 screening using chest X-ray images: (I) Normal vs COVID-19, (II) Pneumonia vs COVID-19 and (III) Non-COVID-19 vs COVID-19. The proposed study was validated using the HM Hospitals COVID-19 dataset for the COVID images and the RSNA Pneumonia Challenge dataset for the Normal and Pneumonia images. The DenseNet-161 architecture was adapted for the automatic screening purposes (de Moura, et al., 2020; de Moura, Novo, & Ortega, 2022).

In the first analysis, we explored intermediate imbalance scenarios in which female and male patients diagnosed with COVID-19 were analysed in different proportions with 10% intervals, ranging from 0% male patients and 100% female patients to 100% male patients and 0% female patients. Thus, we conducted a comprehensive analysis with 11 different configurations for each computational

approach. For each imbalance case, we get a model that is then tested using the remaining images not used during training. Afterwards, we compare the results obtained for each scenario with our baseline (50% female and 50% male).

For the age-related imbalance study, we defined 6 different age ranges: 0-40, 40-50, 50-60, 60-70, 70-80, >=80. For each range, we used only images from patients in that age spectrum for training and then tested it with the remaining images. We analysed the differences between the age group used for training, which acts as our baseline, and all other ages. In addition, to adapt to the quantity of samples we had available from the investigated Normal, Pneumonia, and COVID-19 classes of interest, we attempted to emphasize the older age groups since they experience the disease more severely and require a more thorough diagnosis method.

3 Results and Discussion

Regarding the sex-related imbalance analysis, the precision, recall and F1-score measures were in every experiment in all the approaches above 96%, which is a satisfactory result. As for accuracy, the obtained measures for every experiment for each approach were above 96% with a standard deviation under 10%. There were no extreme peaks in either the accuracy or its standard deviation in none of the approaches, and differences between experiments and approaches are around 5%. Despite these differences, the accuracy remained stable and similar to other approaches. All these satisfactory results, together with the stability observed in all the scenarios considered in each of our approaches, indicate that this factor has not clearly affected the diagnosis offered by our system. Thereby, no influence caused by the sex factor was observed.

Regarding the age-related imbalance analysis, the precision, recall and F1-score measures were in every experiment in all approaches above 96%, which is a satisfactory result. As for accuracy, the obtained results for each approach were above 80% with a standard deviation under 15%. The standard deviation increased as baseline patients got older than 70. Although the closer to the baseline age the tested age range gets, the better accuracies are obtained, these differences are not of great magnitude. It is noteworthy that this worsening is more or less present in all the cases studied, but is more pronounced in the older age groups, which is consistent given that the most critical cases of COVID-19 are more frequent in this group, resulting in a greater variability of pathological affectations in the lungs. Hence, it could justify the presence of this bias.

In conclusion, all the results showed that the proposed methodology and tested approaches provide a robust and reliable analysis to support clinical decision-making.

References

Álvarez-Rodríguez, L., de Moura, J., Novo, J., & Ortega, M. (2022). Does imbalance in chest X-ray datasets produce biased deep learning approaches for COVID-19 screening? *BMC Medical Research Methodology, 22*, 1–17.

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., . . . Mavridis, N. (2020, June). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digital Medicine, 3*. doi:10.1038/s41746-020-0288-5

de Moura, J., Novo, J., & Ortega, M. (2022). Fully automatic deep convolutional approaches for the analysis of Covid-19 using chest X-ray images. *Applied Soft Computing, 115*, 108190.

de Moura, J., Ramos, L., Vidal, P., Cruz, M., Abelairas, L., Castro López, E., . . . Ortega Hortas, M. (2020). Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices. *IEEE Access, 8*, 195594–195607.

Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020, May). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences, 117*, 12592–12594. doi:10.1073/pnas.1919012117

Morís, D. I., de Moura, J., Novo, J., & Ortega, M. (2021). Cycle generative adversarial network approaches to produce novel portable chest x-rays images for covid-19 diagnosis. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (pp. 1060–1064).

Vidal, P. L., de Moura, J., Novo, J., & Ortega, M. (2021). Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19. *Expert Systems with Applications, 173*, 114677. doi:10.1016/j.eswa.2021.114677

WHO Coronavirus (COVID-19) Dashboard. (n.d.). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved August 3, 2022