EasyChair Preprint

№ 13711

Imbalanced Datasets and Bias in Artificial Intelligence: Influence of Sex and Age for COVID-19 Screening

Lorena Álvarez, Joaquim de Moura, Jorge Novo and Marcos Ortega

June 19, 2024

Imbalanced datasets and bias in artificial intelligence: influence of sex and age for COVID-19 screening

Lorena Álvarez-Rodríguez*, Joaquim de Moura*, Jorge Novo*, Marcos Ortega*
* *Grupo VARPA, INIBIC, University of A Coruña, A Coruña, Galicia, Spain*
{lorena.alvarezr, joaquim.demoura, jnovo, mortega}@udc.es

***Abstract*—In this study, we examined eleven imbalance scenarios in which COVID-19 patients were present in varying proportions for the sex analysis and six scenarios in which the age factor was trained using only one particular age range. Three distinct methods for automatically detecting COVID-19 were employed in each study: (I) COVID-19 VS Normal, (II) COVID-19 vs Pneumonia, and (III) Non-COVID-19 VS COVID-19. Two representative public chest X-ray datasets were used to validate the current work, enabling a trustworthy analysis to aid in clinical decision-making. The findings of the sex-related analysis show that this element has a minor impact on the COVID-19 deep learning-based systems, but not enough to significantly degrade the system. Age was shown to be influencing the system more consistently in the age-related study because it was present in every scenario that was taken into consideration.**

***Index Terms*—COVID-19 screening, chest X-rays, deep learning, data analysis**

I. Introduction

Our study was motivated by the COVID-19 outbreak, classified by the World Health Organization (WHO) as a pandemic in March 2020. It resulted in millions of cases and deaths, and the need for quick, accurate, and secure means of detecting and tracking respiratory infections had never been more apparent. To this end, chest X-rays are typically used for examining impacted areas in lungs, but a thorough analysis is required, which should be carried out by a qualified expert.

This lead to many works proposing automatic artificial intelligence systems that allowed a in-depth analysis of these medical images for COVID-19. However, these techniques are influenced by data scarcity, which affects even more widely studied and prevalent diseases like cancer or pneumonia, where public datasets are few and sometimes unbalanced. COVID-19 was an on-going problematic at the time of our study [1], with few small COVID-19 datasets being used along others much different, like pediatric pneumonia samples. There is evidence bias in terms of sex and gender, among other patient characteristics, produces discriminatory results in medical deep learning systems. However, when our work was published in 2022 [1], data scarcity in COVID-19 chest X rays deep learning systems was barely addressed, but no other study, specifically for sex and age, had yet been carried out regardless of the volume of articles and research, the urgency, and the lack of COVID-19 datasets.

II. Materials and methods

We performed a comprehensive analysis of sex and age factors in the COVID-19 datasets [1] by defining 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 and RSNA Pneumonia Challenge datasets. The DenseNet-161 architecture was adapted for automatic screening.

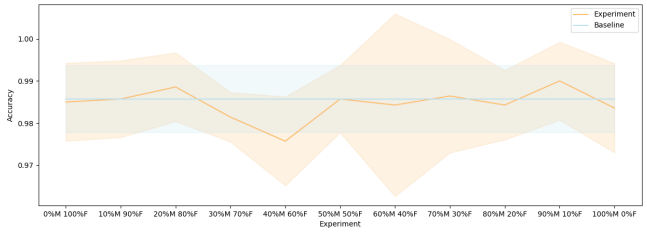
For the sex-related study, we explored intermediate imbalance scenarios in which female and male patients diagnosed with COVID-19 were analysed in different proportions with 10% intervals, ranging from 0% male patients and 100% female patients to 100% male patients and 0% female patients. Thus, we conducted 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 considering during training. Afterwards, we compare the results obtained for each scenario with our baseline (50% female and 50% male). For COVID-19 700 images from 350 female and 350 patients were used, so Normal and Pneumonia classes were equally balanced.

Regarding age, we defined 6 different age ranges: 0-40, 40-50, 50-60, 60-70, 70-80, >=80. For each, we used only images from patients in that age spectrum for training and then tested it with the remaining. 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 every class, we attempted to emphasize the older age groups since they experience the disease more severely and require a more thorough diagnosis method. Samples were downsampled depending on age.

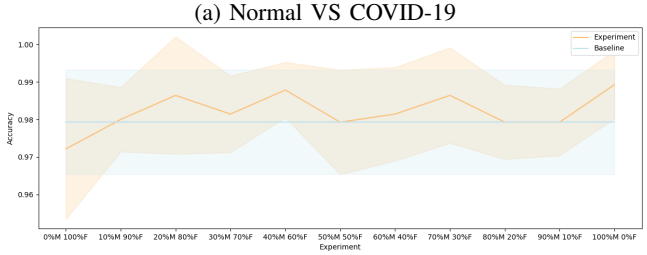
III. Results

For the sex-related analysis, the precision, recall and F1-score measures were in every experiment in all the approaches above 96%. As for accuracy, summarized in Figure 1, 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, it remained stable and similar to other approaches.

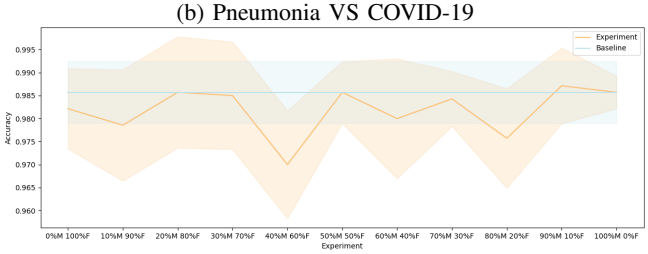
Regarding the age factor, the precision, recall and F1-score measures were in every experiment in all approaches above



(a) Normal VS COVID-19

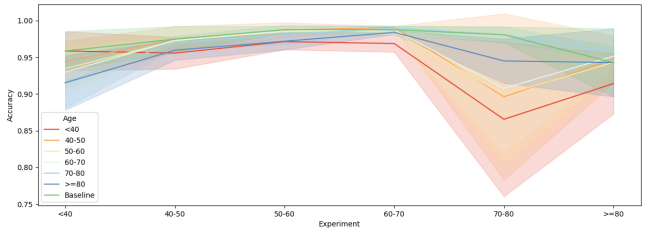


(b) Pneumonia VS COVID-19

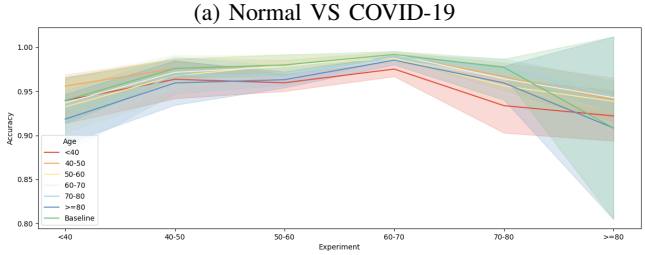


(c) Non-COVID-19 VS COVID-19

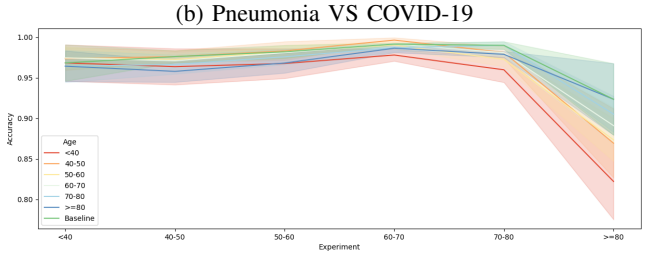
Fig. 1: Comparative analysis of the performance in each imbalance scenario using accuracy for the sex analysis.



(a) Normal VS COVID-19



(b) Pneumonia VS COVID-19



(c) Non-COVID-19 VS COVID-19

Fig. 2: Comparative analysis of the performance in each imbalance scenario using accuracy for the age analysis.

96%. For accuracy summarized in Figure 2, 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 this worsening is present in all cases studied and aggravated for older age groups.

IV. Conclusions

In conclusion, all the results showed our study [1] and tested approaches provide a robust and reliable analysis of possible bias. In terms of sex factor, satisfactory results denote this characteristic has not clearly affected the diagnosis, hence not being observed an influence caused by the sex factor. Although male and female patients may have differentiating features visible in chest X-rays, such as breasts or differences in shape and size, these typically sex-associated features do not influence their COVID-19 diagnosis and do not favor one sex over the other, as they do not interfere with the lung assessment. Regarding age, despite small, some influence was observed with better results the closer to the test age the train data was, being more pronounced in older age groups, which is consistent given most critical cases of COVID-19 are more frequent in this group, resulting in greater variability of pathological affectations in lungs. For example, older patients are usually recognized by the wide range of different damaged ribcages, being these caused by diseases or passing of time. These patients are typically weaker when facing an aggressive disease as COVID-19, so different types of medical equipment, like pathways or thoracostomy tubes, among other cardiac and pulmonary devices, are more present in these X-rays. These elements can appear obscuring lung densities typical of COVID-19 or leading systems to recognize patients more by the irregularity of their X-rays than by the signs of disease they may manifest, both affecting their diagnosis. However, these elements do not appear as frequently in images of younger patients, whose abnormalities are easily observed and linked to COVID-19, because they do not usually have other pathologies that may cause the presence of irregularities in their images.

V. Acknowledgements

Funding by Instituto de Salud Carlos III, Government of Spain (ref. DTS18/00136); Ministerio de Ciencia e Innovación y Universidades, Government of Spain (ref. RTI2018-095894-B-I00); Ministerio de Ciencia e Innovación, Government of Spain (ref. PID2019-108435RB-I00); Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva, grant ref. ED431C 2020/24; postdoctoral grant ref. ED481B 2021/059; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38.

References

[1] L. Álvarez-Rodríguez, J. d. Moura, J. Novo, and M. Ortega, "Does imbalance in chest x-ray datasets produce biased deep learning approaches for covid-19 screening?," *BMC Medical Research Methodology*, vol. 22, no. 1, pp. 1–17, 2022.