

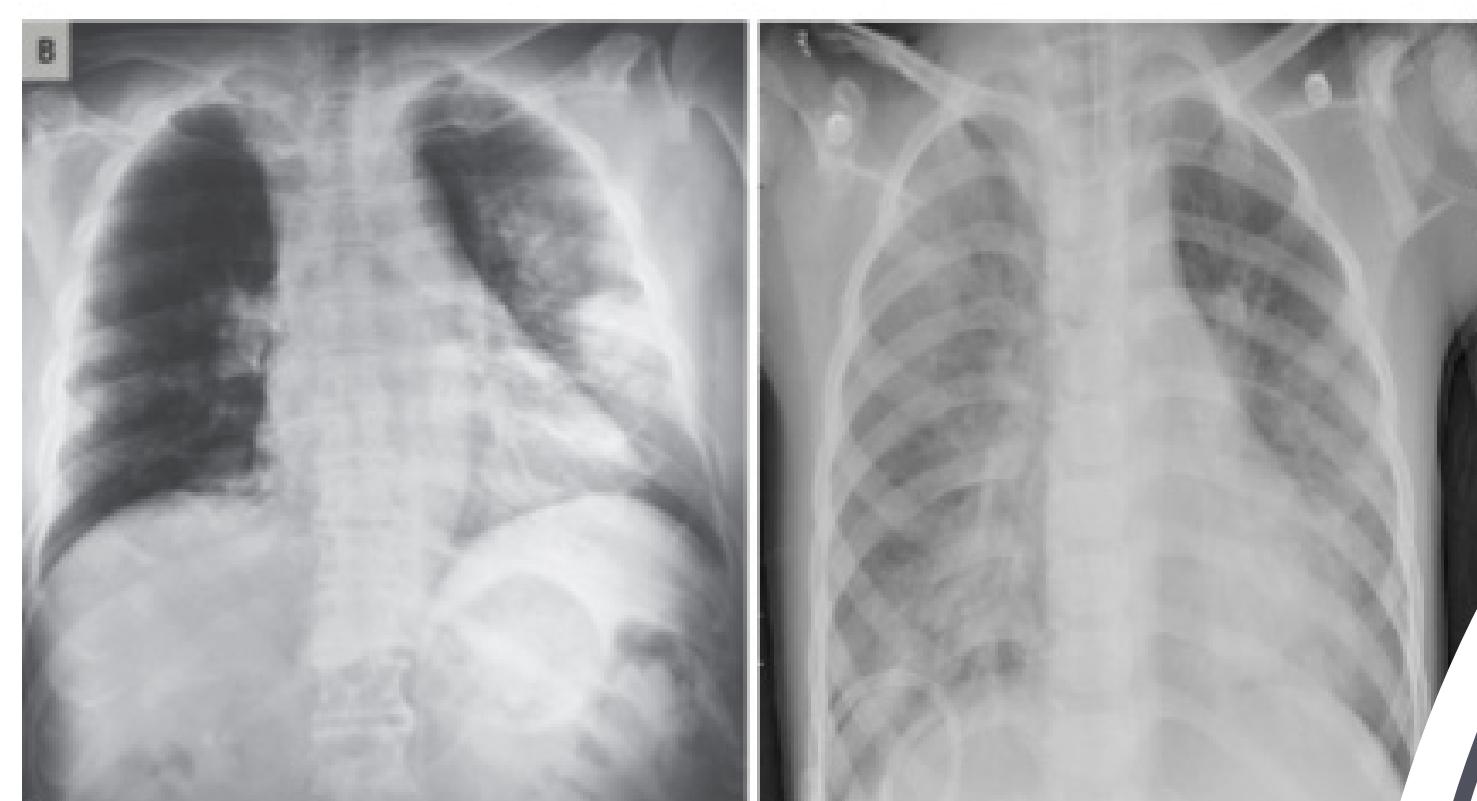
DEALING WITH SCARCE LABELLED DATA: SEMI-SUPERVISED DEEP LEARNING WITH MIX MATCH FOR COVID-19 DETECTION USING CHEST X-RAY IMAGES

S. Calderon-Ramirez¹, R. Giri², S. Yang³, A. Moemeni⁴, M. Umaña⁵, D. Elizondo⁶, J. Torrents⁷, M. Molina-Cabello⁸

Instituto Tecnológico de Costa Rica¹, De Montfort University^{1,2,3,6}, Nottingham University⁴, CCSS-Costa Rica⁵, HP Inc.⁷, Universidad de Málaga⁸

INTRODUCTION

Coronavirus is now a global pandemic, thus the detection of infected subjects in an early, quick and cheap manner is urgent. Deep learning for Covid-19 detection using X-ray images is an attractive approach, however it requires large labelled datasets. To overcome this challenge, a semi-supervised deep learning (SSDL with MixMatch) model using both labelled and unlabelled data is proposed, with the MixMatch architecture to classify chest X-rays into Covid-19, pneumonia and healthy cases. We also introduce a semi-supervised deep learning boost coefficient aimed to ease performance comparison of semi-supervised architectures.



PROPOSED SOLUTION

Use Mix Match for SSDL leveraging unlabelled data with limited labelled observations, and evaluate its scalability with the proposed SSDL boost coefficient. To compare SSDL methods, a measurement including different aspects is important, taking into account the proportion of unlabeled and evaluation data, through the usage of the proposed

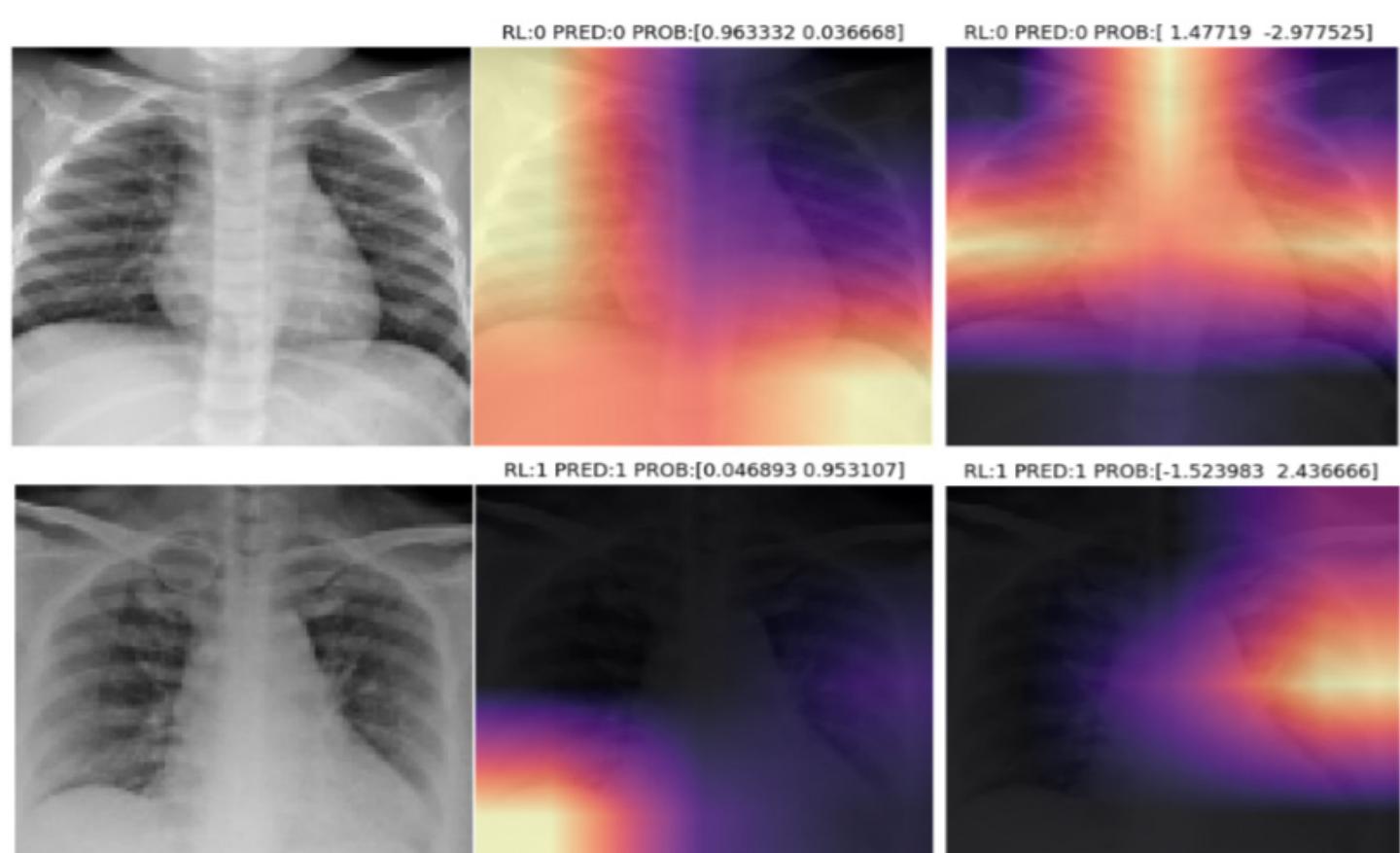
$$\text{Ratio of labeled data: } \rho_{lu} = \frac{n_l}{n_u + n_l}$$

$$\text{Ratio of test data: } \rho_{le} = \frac{n_v}{n_v + n_l}$$

$$\text{SSDL boost coefficient: } \Delta_p = \frac{\bar{a}_{\text{semi-supervised}} - \bar{a}_{\text{supervised}}}{(\rho_{le} + \rho_{lu}) s_{\text{semi-supervised}}}$$

DATA AND EXPERIMENTS

The data used is openly available from Dr. Cohen's repository for COVID-19+ cases [6], and a dataset of Chinese patients for COVID-19- cases. The dataset contains chest X-ray images from around 100 patients, with ages ranging from 27 to 85 years old. For the pneumonia and normal observations, we used the data available in [5]. From such dataset, we selected 4273 observations of pneumonia (including viral and bacterial) and 1583 of normal patients. To avoid a class bias, in most of this work we use an under sampled dataset, containing 102 images for each class. We tested different gamma coefficients with a WideResNet model, to evaluate the impact of unlabeled data. We used 25% of the dataset for testing, and reported the best accuracy along training, implementing early stopping to avoid overfitting.



Heatmaps for the input images (left column), for the supervised model (center), and SSDL model (right)

SSDL WITH MIXMATCH

Mix Match minimizes the following loss function

$$\mathcal{L}(S, w) = \sum_{(x_i, y_i) \in S'_l} \mathcal{L}_l(w, x_i, y_i) + \gamma \sum_{(x_j, \tilde{y}_j) \in \tilde{S}'_u} \mathcal{L}_u(w, x_j, \tilde{y}_j)$$

Where the first labelled term uses the augmented labelled dataset and the second one uses the unlabelled augmented dataset (weighted by gamma), which is obtained using the MixUp augmentation procedure, which creates linear combinations of the data:

$$(S'_l, \tilde{S}'_u) = \Psi_{\text{MixUp}}(S_l, \hat{S}_u, \alpha)$$

For the unlabelled term, MixMatch measures the consistency of the model output and the pseudo-labels using unlabelled augmented data:

$$\mathcal{L}_u(w, x_j, \tilde{y}_j) = \|\tilde{y}_j - f_w(x_j)\|$$

The pseudo-label is calculated as the average label of the model when transforming the input:

$$\hat{y}_j = \frac{1}{K} \sum_{\eta=1}^K f_{\vec{w}}(\Psi^\eta(x_j))$$

Sharpened label has proven to improve the

$$\tilde{y} = \frac{\hat{y}_i^{1/T}}{\sum_j \hat{y}_j^{1/T}}$$

RESULTS AND CONCLUSIONS

We tested different unlabelled term weights, with positive gains when few labels are available of up to 15%, making the approach useful during the beginning of a virus outbreak. Heatmaps in general showed to be more semantically relevant for the semi-supervised model. The proposed semi-supervised accuracy boost coefficient enables the comparison of SSDL methods in practical applications, along a scalability analysis with different labelled/unlabelled data sample sizes. Scalability problems for the tested MixMatch approach were revealed, with less gain using more labels. Future work: analyzing the impact of SSDL in model robustness and uncertainty is important for medical applications.

TABLE I
SEMI-SUPERVISED LEARNING ACCURACY (MEAN AND STD.) USING MIX MATCH (MM) FOR DIFFERENT UNSUPERVISED COEFFICIENTS VS. A FULLY SUPERVISED MODEL (F.S.). ALWAYS $\rho_{lu} = 1$ FOR THE FULLY SUPERVISED MODEL. THE SIXTH COLUMN DENOTES THE CONFIDENCE P-VALUE OF THE ACCURACY DIFFERENCE BETWEEN MIX MATCH AND THE SUPERVISED MODEL.

Number of labels/coefficients	Fully supervised	$\gamma = 1$	$\gamma = 100$	$\gamma = 200$	ES vs. MM ($n = 200$)	$\Delta_p (\gamma = 200)$
25 ($\rho_b = 0.24, \rho_u = 0.11$)	0.683 ± 0.056	0.808 ± 0.053	0.816 ± 0.051	0.826 ± 0.057	$p = 2.36e-04$	7.318
40 ($\rho_b = 0.33, \rho_u = 0.17$)	0.729 ± 0.018	0.828 ± 0.04	0.845 ± 0.048	0.846 ± 0.048	$p = 0.0016$	4.875
50 ($\rho_b = 0.39, \rho_u = 0.21$)	0.785 ± 0.046	0.834 ± 0.038	0.843 ± 0.047	0.843 ± 0.049	$p = 0.0163$	1.972
70 ($\rho_b = 0.47, \rho_u = 0.3$)	0.808 ± 0.046	0.848 ± 0.053	0.861 ± 0.039	0.858 ± 0.041	$p = 0.1155$	1.5838
100 ($\rho_b = 0.56, \rho_u = 0.43$)	0.851 ± 0.049	0.853 ± 0.033	0.856 ± 0.051	0.854 ± 0.047	$p = 0.5194$	0.0648
All-undersampled (229)	0.896 ± 0.035					
All-imbalanced (446)	0.966 ± 0.003					

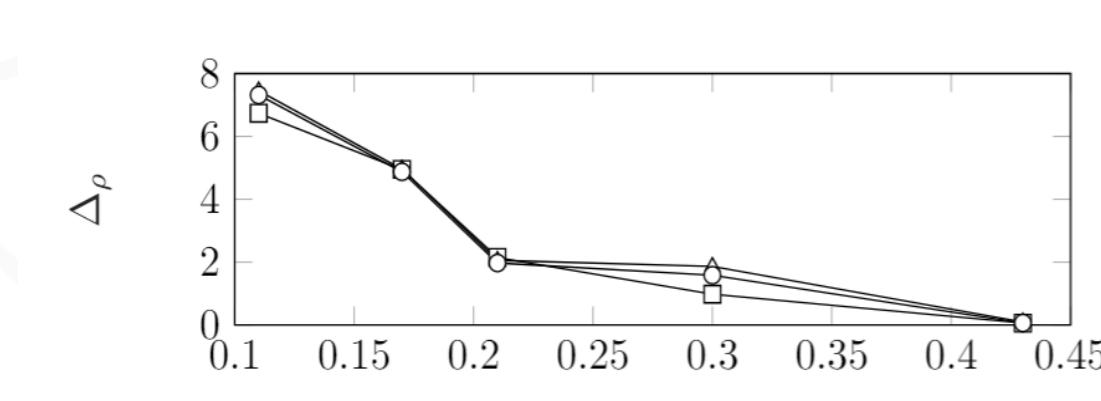


Fig. 4. Scalability curves using Δ_p against the ρ_{lu} . $\gamma = 200$ (triangle), $\gamma = 100$ (circle) and $\gamma = 1$ (square).

SSDL accuracy gains against supervised learning, and scalability curves with different labelled/unlabelled data ratios.

REFERENCES

- [1] Coronavirus - an overview – ScienceDirect Topics.
- [2] COVID-19 DATABASE – SIRMT.
- [3] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Lining Xia. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, page 200642, 2020.
- [4] Indranil Balki, Afsaneh Amirabadi, Jacob Levman, Anna L Martel, Ziga Emersek, Blaz Meden, Angel Garcia-Pedreno, Saul C Ramirez, Dehan Kong, Alan R Moody, et al. Sample-size determination methodologies for machine learning in medical imaging research: A systematic review. Canadian Association of Radiologists Journal, 70(4):344–353, 2019.
- [5] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valen-tim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning.
- [6] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. arXiv 2003.11597, 2020. Available at <https://github.com/ieee8023/covid-chestxray-dataset>
- [7] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In Advances in Neural Information Processing Systems, pages 5050–5060, 2019.

