

Introducing lifelong machine learning in the active tuberculosis classification through chest radiographs

Regina Reis da Costa Alves
Biomedical Engineering Program, COPPE.
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil 21941-901
Email: regina.alves@poli.ufrj.br

Frederico Caetano Jandre de Assis Tavares
Biomedical Engineering Program, COPPE.
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil 21941-901
Email: jandre@peb.ufrj.br

José Manoel de Seixas
Signal Processing Lab, COPPE.
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil 21941-901
Email: seixas@lps.ufrj.br

Anete Trajman
Faculdade de Medicina.
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil 21941-901
Email: atrajman@gmail.com

Abstract—Tuberculosis (TB) is a contagious disease which is among the top 10 causes of death in the world. In order to eliminate the disease by 2050, the treatment of TB infection (TBI) is essential, which requires radiological reports to exclude active tuberculosis. The automatic X-ray classifiers used today are based on models that do not guarantee the retention of knowledge if they need to learn new tasks over time. This work proposes the introduction of the lifelong machine learning (LML) paradigm in automatic X-ray classifiers aimed at helping to diagnose active TB (ATB). Two LML algorithms, Efficient Lifelong Learning Algorithm (ELLA) and Learning without Forgetting (LwF), are applied to the TB and pneumonia classification tasks. The results show that it is possible to keep the performance in both tasks with the LML paradigm.

scars from past diseases, an usual finding especially in places with high prevalence of TB, and also that it is able to learn through experts feedbacks. However, commercial applications and algorithms found in the literature are based on models that do not guarantee the retention of knowledge if they need to learn new tasks over time.

The paradigm of lifelong machine learning (LML) may overcome this difficulty, since the model aims at keeping the efficiency obtained in the tasks previously learned, while using the previously acquired knowledge to have a better performance in new tasks. The learning process can be guided in stages to obtain greater efficiency, through feedback from doctors or by learning diseases that are similar to ATB and that can lead to false positives in the diagnosis of ATB by radiography.

In this work, we test two LML algorithms in learning two tasks: pneumonia and tuberculosis classification. The main goal is to observe whether the models are able to learn a second task while retaining a good performance in the first task with the lifelong learning paradigm.

I. INTRODUCTION

Tuberculosis (TB) is a contagious disease among the top 10 causes of death in the world and the leading cause of death through a single infectious agent until March 2020, surpassed only by COVID-19 in number of daily deaths [1].

In order to reduce the number of cases and deaths by Tuberculosis, the treatment of TB infection (TBI) is essential, as it is the most effective single strategy for the reduction of the number of cases and deaths by 2050 [2]. TBI is a state of persistent immune response to contact with *Mycobacterium tuberculosis*, with no evidence of clinical manifestation of active TB (ATB) [3].

Before beginning TBI treatment, it is necessary to exclude the possibility of the individual presenting ATB [3]. In Brazil, a chest x-ray (CXR) is mandatory to exclude the possibility of the patient having ATB [4]. The need for an expert doctor to issue a radiological report for the candidate for TBI treatment delays the start of the treatment in many primary care settings ([5], [6]).

In order to avoid this source of delay in starting TBI treatment, a computer-aided diagnosis (CAD) may be adopted to classify CXRs. It is useful that the CAD model learns over time to differentiate ATB from other similar diseases or

II. MACHINE LEARNING IN TUBERCULOSIS

The use of machine learning (ML) in TB is not limited to the classification of CXRs. Examples of other applications for TB classification include analysis of blood samples ([7], [8], [9], [10]), genetics [11], exhaled air ([12], [13]), intracellular microRNAs [14], clinical data ([15], [16]) and computed tomography [17].

Among the ML techniques used for TB screening, neural networks, Decision Trees, Support Vector Machines (SVM), Bayes Naïve Classifier, Genetic Algorithms and Fuzzy Logic are often used [18].

For the classification of TB using CXR, the most common approach is the use of convolutional neural networks (CNN), since they have become the preferred technique for analyzing medical images [18]. The sets of parameters present in the

convolutional layers extract visual features from the image, while the parameters of the fully-connected layers are responsible for the classification.

Hwang *et al.* [19] define a CNN based on the architecture of *AlexNet* [20]. The used datasets were from Shenzhen, Montgomery [21] and the Korean Institute of Tuberculosis (KIT). They have obtained an average accuracy of 90.3%.

Lakhani *et al.* [22] apply two CNNs, AlexNet and GoogLeNet, to the Montgomery, Shenzhen, Belarus and the Thomas Jefferson University Hospital datasets. Images that had a classification disagreement between AlexNet and GoogLeNet were sent to a radiologist to be classified, which improved the classification performance, resulting in a sensitivity of 97.3% and a specificity of 100%.

Becker *et al.* [23] apply a CNN for the identification of patterns that appear in CXR of patients with ATB. The identification of areas with pathologies in the CXR obtained a ROC of 0.82.

Pasa *et al.* [24] present a CNN optimized for the TB classification problem through CXR. After the network was trained for classification, saliency maps and gradient maps of class activation (grad-CAM) were generated. The obtained accuracy was 79% for Montgomery, 84.4% for Shenzhen and 86.2% for the both combined.

Heo *et al.* [25] use CNN to detect ATB in CXR with and without using patient demographic variables. The networks VGG19, InceptionV3, ResNet50, DenseNet121 and Inception-ResNetV2 were used. The best results in terms of AUC were obtained with VGG19: 0.9075 for the case without demographic variables and 0.9213 for the case with demographic variables.

Hwang *et al.* [26] develop an algorithm to classify ATB, pneumonia, malignant lung neoplasms and pneumothorax. A deep CNN was used with 5 classifiers in parallel, with 4 specialized classifiers for each disease and the fifth classifying any anomaly, encompassing features from the 4 diseases. The algorithm obtained an accuracy of 98.3%.

There are also several works applying SVM to the CXRs classification problem. SVM is a machine learning algorithm which maps the data in a feature space of higher dimension and defines a hyperplane in this space, which separates the data into classes with the maximum margin. With this, a non-linear decision boundary is obtained in the original data space for class separation [27]. The main differences between these works are related to the extracted features for the classification with SVM.

Chauhan *et al.* [28] extracted features with the GLCM method, Gabor filters, Gist, Histogram of Oriented Gradients (HOG) and Pyramid Histogram of Oriented Gradients (PHOG). Udayakumar *et al.* [29] extracted the lung region using the graph cut segmentation method. In the extracted segment, features related to content-based image recovery (CBIR) and features inspired by object detection are computed. Kamble *et al.* [30] used CBIR and the calculation of standard deviation, mean, kurtosis, skewness and entropy of the image histograms. Chandra *et al.* [31] proposed an algorithm in

which features are extracted in a hierarchical organization. At level 1, geometric features such as shape, size and eccentricity are extracted and at level 2, first-order statistical features such as energy, entropy, contrast and correlations are extracted.

Although these algorithms found in the literature have good performance in the task which they were designed for, they did not explore the possibility of retaining knowledge if they need to learn new tasks over time. This is a useful feature considering that there are several similar diseases, as well as scars from past diseases, which can lead to errors in classification and the model can learn to differentiate them over time through experts feedbacks.

III. LIFELONG MACHINE LEARNING

A definition of lifelong learning is found in [32] as: “Lifelong Machine Learning, or LML, considers systems that can learn many tasks over a lifetime from one or more domains. They efficiently and effectively retain the knowledge they have learned and use that knowledge to more efficiently and effectively learn new tasks.”

There are three main elements for this process [32]: (1) the retention of previously learned knowledge; (2) the selective knowledge transfer to learn new tasks; and (3) an approach that ensures effective and efficient interaction between the elements retained and transferred between tasks. Failure to observe the first point highlighted is known in the literature as catastrophic forgetting.

In neural networks, catastrophic forgetting happens due to the plasticity-stability dilemma [33]. If the net is too plastic, it will tend to overwrite its weights while learning new tasks and start to perform poorly on the previous ones, but, if it is very stable, it will not be able to learn new tasks.

Different strategies can be adopted to implement the LL:

- Network growth strategies: based on adding parameters to the network in order to allow the learning of new tasks and, at the same time, retain previous knowledge through the old parameters [34], [35], [36].
- Selective plasticity and neuromodulation strategies: methods that allow a subgroup of parameters to be adjusted to the new task while another subgroup is almost or completely frozen [37], [38], [39], [40].
- Generative strategies: inspired by the existence of a dual memory system based on the hippocampus and neocortex in the human brain. The hippocampus encodes recent experiences and, acting as a generative model, generates multiple repetitions of the experience during sleep and the memory is consolidated in the neocortex [41], [42], [43].
- Memorization strategies: depend on the storage of all or part of data from old tasks to control the preservation of knowledge [44], [45], [46], [47].
- Penalization strategies: add a penalty term to the loss function in order to reduce deviations in the classification performance of previous tasks [48], [49], [50].

- Latent components strategies: consider knowledge sharing through latent components shared between different tasks [51], [52], [53], [54], [55], [56], [57].

There are few studies that apply LML in medicine and, specifically, in the CXR classification. The work developed by Kim *et al.* [58] deals with a scenario in which data are spread across multiple hospitals and a machine learning model must be trained sequentially. The model was tested on the CIFAR-10/100 dataset and on the CXR dataset for Tuberculosis at the Korean Tuberculosis Institute [19]. Patra *et al.* [59] have developed a study based on the diagnosis of congenital heart disease in fetuses, which in general have their diagnosis based on small regions of interest in the images. The diagnosis is made based on classes of interest, which are important anatomical structures for this activity.

IV. APPLICATION OF TWO LML ALGORITHMS TO DISEASE CLASSIFICATION TESTS

In order to make first experiments applying LML algorithms to diseases classification through chest X-rays images, two distinct tasks were considered: the classification of pneumonia and tuberculosis. For pneumonia, data from the Stanford public dataset [60] were used, and for tuberculosis, data from the Shenzhen public dataset [21] were used.

The Stanford dataset has 14 classes, including Normal, Pneumonia and Tuberculosis, and was undersampled according to the following criteria:

- Only 2 classes: “Normal” and “Pneumonia”.
- Images present in the reduced set available for download (11 GB, while the full set has 439 GB).
- Front images and AP view.
- Similar amount of images in both classes.

The whole Shenzhen dataset was used. The number of images per class that were used in the studies is in Table I.

TABLE I
NUMBER OF IMAGES PER CLASS.

Class	Quantity
Non-Tuberculosis (Shenzhen)	326
Tuberculosis (Shenzhen)	336
Normal (Stanford)	1070
Pneumonia (Stanford)	1065

For both algorithms, two experiments were made to learn the tasks. In the first experiment, the first task was to classify pneumonia and the second one was to classify tuberculosis, while in the second experiment, task order was reversed. These algorithms are not capable of autonomously detecting new tasks, and the tasks must be explicitly presented.

The first LML algorithm chosen for testing was the Efficient Lifelong Learning Algorithm (ELLA) [51]. This algorithm is among the relevant pioneers in LML. For classification problems, as in this work, ELLA is based on a logistic regression model. The code used in this study was made available by the authors.

The algorithm is based on maintaining a library of latent components as a shared knowledge base for the specific

models associated with each task. Let $L \subseteq R^{d \times k}$ be the base of latent components and d the dimension of inputs x . It is assumed that the vector θ^t , corresponding to the parameters of each task t , is a linear combination of the components in L whose weights are given by the vector $s^t \in R^k$. That is, for each task t :

$$\theta^t = Ls^t \quad (1)$$

We are left with the following Objective Function to be minimized, depending on the L and the s^t of each task:

$$\frac{1}{N} \sum_{t=1}^N \min_{s^t} \left\{ \frac{1}{n_t} \sum_{i=1}^{n_t} \mathcal{L}(f(x_i^t; \theta^t), y_i^t) + \mu \|s^t\|_1 \right\} + \lambda \|L\|_F^2 \quad (2)$$

Where:

$$f(x_i^t; \theta^t) = \frac{1}{1 + e^{-(\theta^t)^T x_i^t}} \quad (3)$$

The mechanism to control the forgetting of past tasks is the term in the Objective Function which penalizes errors in classification with respect to all tasks. In order to remove the dependence from all of the previous training data, the authors approximate Equation 2 using the second-order Taylor expansion around $\theta^{(t)} = \underset{\theta}{\operatorname{argmin}} \frac{1}{n_t} \sum_{i=1}^{n_t} \mathcal{L}(f(x_i^{(t)}; \theta), y_i^{(t)})$, which is an optimal predictor learned on only the training data for task t .

When a new task arrives, it uses the stored latent components and improves the base L with the knowledge acquired in the new task. This means that, over time, the performance of past tasks can also be improved. An assumption for ELLA is that inputs from all tasks must come from the same space R^d . As the model is based on a logistic regression, all tasks must have only two classes.

In the experiments, the inputs were the x-ray images, which were pre-processed using the Principal Component Analysis for the retention of different percentages of variance (85%, 90% and 95%), with the aim of reducing the dimensionality of the problem. The values of hyperparameters $\mu = 1.0$ and $\lambda = 10^{-5}$ were kept equal to those defined by the authors in the code provided by them. The number of latent components $k = 8$ was chosen after a gridsearch over values of k from 1 to 15 for task 1 = TB classification and task 2 = pneumonia classification. Note in Figure 1 that, with $k = 8$, there is a significant improvement in sensitivity and specificity specially for task 2 with 95% of variance retention, while the other indices are less affected by the number of latent components.

The cross-validation method, specifically the k-folds technique with $k=10$, was used to estimate the uncertainty in the results due to the fact that we are using a limited sample. The results obtained for both experiments in the test sets are presented in Table II and III. They correspond to the confidence interval at a confidence level of 95%.

The second LML algorithm studied herein is the *Learning without Forgetting* (LwF) [49]. This algorithm is based on a CNN, which has become the preferred technique for analyzing medical images [18]. At the same time, LwF is a relevant

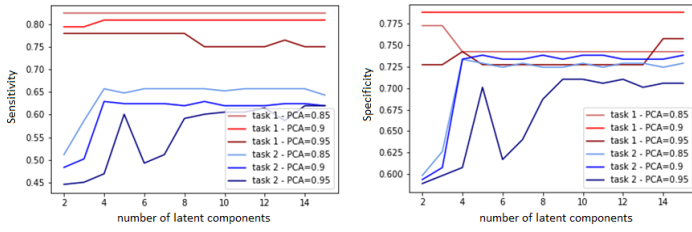


Fig. 1. Sensitivity and specificity as a function of number of latent components.

TABLE II
RESULTS FOR ELLA. FIRST TASK WAS TO CLASSIFY PNEUMONIA AND SECOND TASK WAS TO CLASSIFY TUBERCULOSIS.

Index	PCA = 85%	PCA = 90%	PCA = 95%
Sensitivity (pneumonia)	73.0 ±12.4%	74.1 ±11.9%	70.8 ±10.8%
Specificity (pneumonia)	70.3 ±9.5%	67.9 ±10.0%	66.8 ±7.0%
Error (pneumonia)	28.4 ±7.5%	29.0 ±8.3%	31.2 ±7.7%
Sensitivity (tuberculosis)	78.9 ±14.4%	77.4 ±11.4%	76.5 ±12.6%
Specificity (tuberculosis)	86.4 ±13.6%	81.5 ±19.1%	82.5 ±15.6%
Error (tuberculosis)	18.3 ±8.4%	20.6 ±7.7%	20.6 ±7.4%

algorithm in the literature and is based on a simple strategy to control catastrophic forgetting, by penalizing the Objective Function. Another desirable feature is that this algorithm does not require data storage from previous tasks.

The LwF algorithm is based specifically on the *Alexnet* network, so the network architecture was prefixed. The network has parameters θ_s shared among all tasks, corresponding to all layers of the network except the last one, and parameters θ_o which are specific to each task, corresponding to the last layer of the network.

When a dataset related to a new task arrives, it is presented to the network and the outputs y_o corresponding to each of the previous tasks last layer are recorded. Next, parameters θ_n related to the new task are added to the output layer, connected to all neurons in the previous layer and initialized with random values.

The network is trained to minimize classification errors for all tasks through a modified loss function. For older tasks, the classification error is the deviation of the updated network outputs from the saved y_o outputs. During training, first the parameters θ_o and θ_s are fixed and only the parameters θ_n are adjusted until it achieves a stop criterion (warm-up phase). Then all parameters are adjusted up to a stop criterion (adjustment phase). The authors used the number of epochs as the stop criterion, which was also adopted in the tests made in this work.

Like ELLA, for both tasks, the inputs must come from the same space R^d for LwF; in this case, for both tasks, the inputs are 224x224 RGB images. Also note that each task may have a different number of classes, as the last layer of the network is task specific.

The LwF modified loss function, or Objective Function (OF) to be minimized, is formulated as:

TABLE III
RESULTS FOR ELLA. FIRST TASK WAS TO CLASSIFY TUBERCULOSIS AND SECOND TASK WAS TO CLASSIFY PNEUMONIA.

Index	PCA = 85%	PCA = 90%	PCA = 95%
Sensitivity (tuberculosis)	79.5 ±14.2%	77.1 ±9.2%	77.1 ±15.0%
Specificity (tuberculosis)	84.6 ±15.1%	81.9 ±14.7%	80.6 ±16.9%
Error (tuberculosis)	18.0 ±7.6%	20.5 ±6.8%	21.2 ±10.4%
Sensitivity (pneumonia)	73.0 ±12.4%	72.7 ±11.5%	70.0 ±10.5%
Specificity (pneumonia)	70.1 ±8.4%	68.1 ±8.5%	66.6 ±14.6%
Error (pneumonia)	28.5 ±7.1%	29.6 ±7.9%	31.7 ±10.9%

$$OF = \lambda_o \mathcal{L}_{old}(y_o, \hat{y}_o) + \mathcal{L}_{new}(y_n, \hat{y}_n) + \mathcal{R}(\hat{\theta}_s, \hat{\theta}_o, \hat{\theta}_n) \quad (4)$$

Where:

$$\mathcal{L}_{old}(y_o, \hat{y}_o) = -H(y'_o, \hat{y}'_o) = -\sum_{i=1}^l y'_o{}^{(i)} \log \hat{y}'_o{}^{(i)} \quad (5)$$

$$\mathcal{L}_{new}(y_n, \hat{y}_n) = -y_n \cdot \log \hat{y}_n \quad (6)$$

$$y'_o{}^{(i)} = \frac{(y_o^{(i)})^{1/T}}{\sum_{j=1}^l (y_o^{(j)})^{1/T}} \quad (7)$$

$$\hat{y}'_o{}^{(i)} = \frac{(\hat{y}_o^{(i)})^{1/T}}{\sum_{j=1}^l (\hat{y}_o^{(j)})^{1/T}} \quad (8)$$

And:

- $\mathcal{R}(\hat{\theta}_s, \hat{\theta}_o, \hat{\theta}_n)$ is a regularization term;
- λ_o is a loss balance weight; the higher the value, the more importance is given to the performance in previous tasks, to the detriment of the new task;
- \hat{y}_n is the last layer output corresponding to the new task;
- y_n is the vector of labels for the new task;
- $\hat{y}_o^{(i)}$ is the current last layer output corresponding to the previous task i ;
- $y_o^{(i)}$ is the recorded last layer output corresponding to the previous task i ;
- 1 is the number of labels;
- $y'_o{}^{(i)}$ and $\hat{y}'_o{}^{(i)}$ are modified versions of the current and recorded output;
- T is a parameter that controls the weight given to output values in the modified version.

To conduct the experiments, both Shenzhen and Stanford datasets were randomly divided, while preserving the proportion of images per class, in two sets: training (80%) and validation (20%). The algorithm implementation was developed by the authors, in Matlab [61]. In this implementation, the first task must be trained outside the main program, since the algorithm already loads the AlexNet network pre-trained with the first task. The hyperparameters during training of the first task were: learning rate of 0.001, 20 epochs and batch size of 20.

Figure 2 shows the classification error evolution with the epochs, in the training set and in the validation set. For both

networks, we chose the set of parameters which lead to the local minimum error in the validation set, before it starts to increase. The stop points are surrounded by a black circle on both graphs.

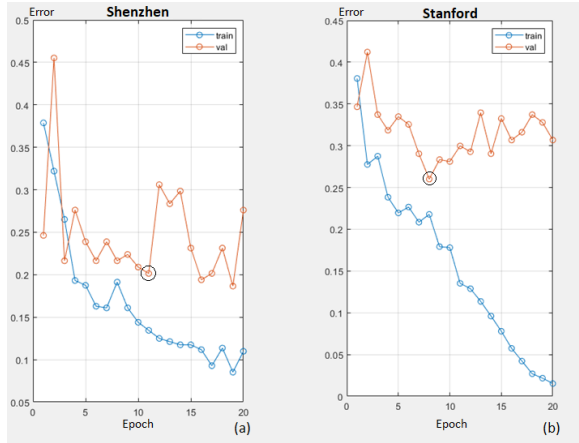


Fig. 2. Evolution of error as a function of time for the Shenzhen (a) and Stanford (b) datasets.

For the learning of the second task, all the hyperparameters were the same that the authors used in their experiments: learning rate of 0.001, 72 epochs in the warm-up phase, 72 epochs in the adjustment phase and batch size of 20.

Figures 3 and 4 show the error evolution with the epochs for Experiment 1 and Experiment 2, respectively.

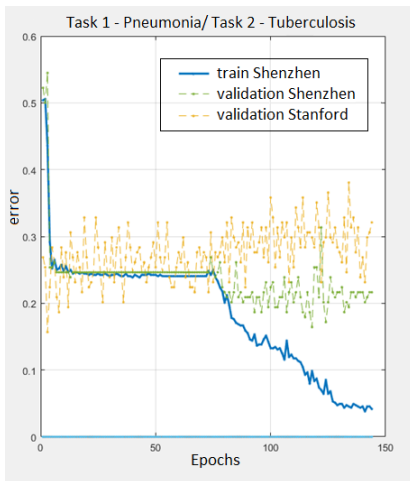


Fig. 3. Error evolution with the epochs. First task was to classify pneumonia and second task was to classify tuberculosis.

In experiment 1, in the warm-up phase, the network quickly reaches an error of approximately 25% in the new task (classification of pneumonia with the Stanford dataset). When all weights in the network are adjusted, the error in the new task validation set starts to oscillate in a lower value ($\sim 20\%$) and error from the previous task (classification of TB with Shenzhen base) starts to oscillate at a higher value ($\sim 30\%$). This behavior is expected, since the shared parameters were

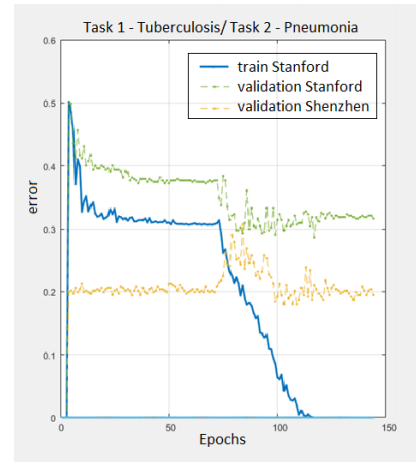


Fig. 4. Error evolution with the epochs. First task was to classify pneumonia and second task was to classify tuberculosis.

adequate for task 1 and started to be adjusted taking into account both tasks. However, it is expected that the decrease in the performance of task 1 would be greater if there was no mechanism for the preservation of knowledge.

In experiment 2, when all weights in the network are adjusted, the error in the previous task validation set (pneumonia classification with the Stanford dataset) considerably drops to around 30%. The error in the first task (Shenzhen) starts increasing, however, later it returns to the previous level ($\sim 20\%$). This shows the effect of the knowledge retention mechanism implemented in the algorithm. Also, note that both experiments ended up with similar performance in the tasks, with an error of approximately 20% for ATB classification and approximately 30% for pneumonia classification.

V. DISCUSSION AND CONCLUSION

This work presented an application of two lifelong machine learning algorithms to the pneumonia and tuberculosis classifications through chest x-ray images. The results show that it is possible to keep the performance in both pneumonia and TB classification with the LML paradigm.

Although one could propose two independent models, one for each task, with similar performance, this paradigm is potentially scalable for more tasks without the need to add a large number of new parameters, as tasks share several parameters. This feature may prove useful not only to reduce size of the model (possibly speeding up its operation and reducing storage demands) but also to share knowledge between tasks and allow a continuous learning and improvement of models.

Proposals for future work include hyperparameter optimization in LwF and ELLA, in order to allow a fair comparison between them and other models in literature, tests with more than two tasks and studies about model explainability. They also include the development of a lifelong machine learning model to learn continuously from experts feedback and applied to a dataset of candidates for TBI treatment. This model may be useful in reducing one of the sources of delay in starting

TBI treatment, which is the lack of expert doctors to issue a radiological report to exclude the possibility of the individual presenting ATB.

VI. ACKNOWLEDGMENTS

The authors would like to thank FAPERJ and CNPq for supporting this work.

This work was carried out with the support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001.

REFERENCES

- [1] D. R. Silva, F. C. d. Q. Mello, L. D'Ambrosio, R. Centis, M. P. Dalcolmo, and . Migliori, "Tuberculose e COVID-19, o novo dueto maldito: quais as diferenças entre Brasil e Europa?" vol. 47, no. 2, pp. e20210044–e20210044, 2021.
- [2] C. Dye, P. Glaziou, K. Floyd, and M. Raviglione, "Prospects for Tuberculosis Elimination," *Annual Review of Public Health*, vol. 34, no. 1, pp. 271–286, 2013, eprint: <https://doi.org/10.1146/annurev-publhealth-031912-114431>. [Online]. Available: <https://doi.org/10.1146/annurev-publhealth-031912-114431>
- [3] "Latent tuberculosis infection: updated and consolidated guidelines for programmatic management." 2018.
- [4] "MANUAL DE RECOMENDAÇÕES PARA O CONTROLE DA TUBERCULOSE NO BRASIL." [Online]. Available: <http://www.aids.gov.br/pt-br/pub/2019/manual-de-recomendacoes-para-o-controle-da-tuberculose-no-brasil>
- [5] J. Melendez, C. I. Sánchez, R. H. H. M. Philipsen, P. Maduskar, R. Dawson, G. Theron, K. Dheda, and B. van Ginneken, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," *Scientific Reports*, vol. 6, Apr. 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4850474/>
- [6] A. H. V. Hoog, H. K. Meme, H. van Deutekom, A. M. Mithika, C. Olunga, F. Onyino, and M. W. Borgdorff, "High sensitivity of chest x-ray reading by clinical officers in a tuberculosis prevalence survey," *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease*, vol. 15, no. 10, pp. 1308–1314, Oct. 2011.
- [7] B. Kaewseekhao, N. Nuntawong, P. Eiamchai, S. Roytrakul, W. Reechaipichitkul, and K. Faksri, "Diagnosis of active tuberculosis and latent tuberculosis infection based on Raman spectroscopy and surface-enhanced Raman spectroscopy," *Tuberculosis*, vol. 121, p. 101916, Mar. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1472979219304627>
- [8] R. Ahmad, L. Xie, M. Pyle, M. F. Suarez, T. Broger, D. Steinberg, S. M. Ame, M. G. Lucero, M. J. Szucs, M. MacMullan, F. S. Berven, A. Dutta, D. M. Sanvictores, V. L. Tallo, R. Bencher, D. P. Eisinger, U. Dhingra, S. Deb, S. M. Ali, S. Mehta, W. W. Fawzi, I. D. Riley, S. Sazawal, Z. Premji, R. Black, C. J. L. Murray, B. Rodriguez, S. A. Carr, D. R. Walt, and M. A. Gillette, "A rapid triage test for active pulmonary tuberculosis in adult patients with persistent cough," *Science Translational Medicine*, vol. 11, no. 515, Oct. 2019, publisher: American Association for the Advancement of Science Section: Research Article. [Online]. Available: <https://stm.sciencemag.org/content/11/515/eaaw8287>
- [9] S. Khan, R. Ullah, S. Shahzad, N. Anbreen, M. Bilal, and A. Khan, "Analysis of tuberculosis disease through Raman spectroscopy and machine learning," *Photodiagnosis and Photodynamic Therapy*, vol. 24, pp. 286–291, Dec. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1572100018302539>
- [10] J. K. Roe, N. Thomas, E. Gil, K. Best, E. Tsaliki, S. Morris-Jones, S. Stafford, N. Simpson, K. D. Witt, B. Chain, R. F. Miller, A. Martineau, and M. Noursadeghi, "Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis," *JCI insight*, vol. 1, no. 16, p. e87238, Oct. 2016.
- [11] S. Leong, Y. Zhao, R. Ribeiro-Rodrigues, E. C. Jones-López, C. Acuña-Villaorduña, P. M. Rodrigues, M. Palaci, D. Alland, R. Dietze, J. J. Ellner, W. E. Johnson, and P. Salgame, "Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB," *Tuberculosis*, vol. 120, p. 101898, Jan. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1472979219303907>
- [12] A. H. J. Kolk, J. J. B. N. van Berkel, M. M. Claassens, E. Walters, S. Kuijper, J. W. Dallinga, and F. J. van Schooten, "Breath analysis as a potential diagnostic tool for tuberculosis," *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease*, vol. 16, no. 6, pp. 777–782, Jun. 2012.
- [13] M. Beccaria, C. Bobak, B. Maitshotlo, T. R. Mellors, G. Purcaro, F. A. Franchina, C. A. Rees, M. Nasir, A. Black, and J. E. Hill, "Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques," *Journal of Breath Research*, vol. 13, no. 1, p. 016005, 2018.
- [14] F. J. Duffy, E. Thompson, K. Downing, S. Suliman, H. Mayanja-Kizza, W. H. Boom, B. Thiel, J. Weiner Iii, S. H. E. Kaufmann, D. Dover, D. L. Tabb, H. M. Dockrell, T. H. M. Ottenhoff, G. Tromp, T. J. Scriba, D. E. Zak, G. Walzl, and GC6-74 Consortium, "A Serum Circulating miRNA Signature for Short-Term Risk of Progression to Active Tuberculosis Among Household Contacts," *Frontiers in Immunology*, vol. 9, p. 661, 2018.
- [15] F. Aguiar, J. Pinto, R. Torres, J. Seixas, and F. Mello, "Artificial neural network (ANN) for prediction of pulmonary tuberculosis in hospitalized patients," *European Respiratory Journal*, vol. 42, no. Suppl 57, Sep. 2013, publisher: European Respiratory Society Section: 10.2 Tuberculosis. [Online]. Available: <https://erj.ersjournals.com/content/42/Suppl57/P4469>
- [16] C. Li, L. Hou, B. Y. Sharma, H. Li, C. Chen, Y. Li, X. Zhao, H. Huang, Z. Cai, and H. Chen, "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, Jan. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169260717309288>
- [17] Z. Xu, U. Bagci, A. Kubler, B. Luna, S. Jain, W. R. Bishai, and D. J. Mollura, "Computer-aided detection and quantification of cavitary tuberculosis from CT scans," *Medical Physics*, vol. 40, no. 11, p. 113701, Nov. 2013.
- [18] S. Meraj, R. Yaakob, A. Azman, S. N. M. Rum, and A. Nazri, "Artificial Intelligence in Diagnosing Tuberculosis: A Review," 2019.
- [19] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," Mar. 2016, p. 97852W.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: <https://doi.org/10.1145/3065386>
- [21] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, Dec. 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256233/>
- [22] P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Apr. 2017, publisher: Radiological Society of North America. [Online]. Available: <https://pubs.rsna.org/doi/full/10.1148/radiol.2017162326>
- [23] A. S. Becker, C. Blüthgen, V. D. Phi van, C. Sekaggya-Wiltshire, B. Castelnovo, A. Kambugu, J. Fehr, and T. Frauenfelder, "Detection of tuberculosis patterns in digital photographs of chest X-ray images using Deep Learning: feasibility study," *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease*, vol. 22, no. 3, pp. 328–335, 2018.
- [24] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization," *Scientific Reports*, vol. 9, no. 1, p. 6268, Apr. 2019, number: 1 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41598-019-42557-4>

- [25] S.-J. Heo, Y. Kim, S. Yun, S.-S. Lim, J. Kim, C.-M. Nam, E.-C. Park, I. Jung, and J.-H. Yoon, "Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, 2019.
- [26] E. J. Hwang, S. Park, K.-N. Jin, J. I. Kim, S. Y. Choi, J. H. Lee, J. M. Goo, J. Aum, J.-J. Yim, J. G. Cohen, G. R. Ferretti, and C. M. Park, "Development and Validation of a Deep Learning–Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs," *JAMA Network Open*, vol. 2, no. 3, pp. e191095–e191095, Mar. 2019, publisher: American Medical Association. [Online]. Available: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2728630>
- [27] B. Scholkopf, A. J. Smola, and B. Schalkopf, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, Mass: MIT Press, Dec. 2001.
- [28] A. Chauhan, D. Chauhan, and C. Rout, "Role of Gist and PHOG Features in Computer-Aided Diagnosis of Tuberculosis without Segmentation," *PLOS ONE*, vol. 9, no. 11, p. e112980, Nov. 2014, publisher: Public Library of Science. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112980>
- [29] E. Udayakumar, S. Santhi, and P. Vetrivelan, "TB screening using SVM and CBC techniques," 2017. [Online]. Available: <https://www.alliedacademies.org/articles/tb-screening-using-svm-and-cbc-techniques.html>
- [30] M. P. A. Kamble, M. V. V. Anagire, and M. S. N. Chamtagoudar, "CXR Tuberculosis Detection Using MATLAB Image Processing," 2016. [Online]. Available: [/paper/CXR-Tuberculosis-Detection-Using-MATLAB-Image-Kamble-Anagire/55037365b53b1debe5132cc7261ebcf8a75ed98d](https://paperkit.net/paper/CXR-Tuberculosis-Detection-Using-MATLAB-Image-Kamble-Anagire/55037365b53b1debe5132cc7261ebcf8a75ed98d)
- [31] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme," *Expert Systems with Applications*, vol. 158, p. 113514, Nov. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420303389>
- [32] D. L. Silver, Q. Yang, and L. Li, "Lifelong Machine Learning Systems: Beyond Learning Algorithms," in *2013 AAAI Spring Symposium Series*, Mar. 2013. [Online]. Available: <https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5802>
- [33] W. C. Abraham and A. Robins, "Memory retention – the synaptic stability versus plasticity dilemma," *Trends in Neurosciences*, vol. 28, no. 2, pp. 73–78, Feb. 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0166223604003704>
- [34] H. Fayek, L. Cavedon, and H. Wu, "Progressive learning: A deep learning framework for continual learning," *Neural Networks*, vol. 128, pp. 345–357, 2020.
- [35] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive Neural Networks," *arXiv:1606.04671 [cs]*, Sep. 2016, arXiv: 1606.04671. [Online]. Available: <http://arxiv.org/abs/1606.04671>
- [36] J. Yoon, E. Yang, J. Lee, and S. J. Hwang, "Lifelong Learning with Dynamically Expandable Networks," *arXiv:1708.01547 [cs]*, Jun. 2018, arXiv: 1708.01547. [Online]. Available: <http://arxiv.org/abs/1708.01547>
- [37] S. Beaulieu, L. Frati, J. Miconi, J. Lehman, K. Stanley, J. Clune, and N. Cheney, "Learning to continually learn," *Frontiers in Artificial Intelligence and Applications*, vol. 325, pp. 992–1001, 2020.
- [38] E. Ben-Iwhiwhu, P. Ladosz, J. Dick, W.-H. Chen, P. Pilly, and A. Soltoggio, "Evolving inborn knowledge for fast adaptation in dynamic POMDP problems," 2020, pp. 280–288.
- [39] P. Sanda, S. Skorheim, and M. Bazhenov, "Multi-layer network utilizing rewarded spike time dependent plasticity to learn a foraging task," *PLoS computational biology*, vol. 13, no. 9, p. e1005705, Sep. 2017.
- [40] J. Allred and K. Roy, "Controlled Forgetting: Targeted Stimulation and Dopaminergic Plasticity Modulation for Unsupervised Lifelong Learning in Spiking Neural Networks," *Frontiers in Neuroscience*, vol. 14, 2020.
- [41] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual Learning with Deep Generative Replay," *arXiv:1705.08690 [cs]*, Dec. 2017, arXiv: 1705.08690. [Online]. Available: <http://arxiv.org/abs/1705.08690>
- [42] R. Kemker and C. Kanan, "FearNet: Brain-Inspired Model for Incremental Learning," Nov. 2017. [Online]. Available: <https://arxiv.org/abs/1711.10563v2>
- [43] G. I. Parisi, J. Tani, C. Weber, and S. Wermter, "Lifelong learning of human actions with deep neural network self-organization," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 96, pp. 137–149, Dec. 2017.
- [44] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental Classifier and Representation Learning," *arXiv:1611.07725 [cs, stat]*, Apr. 2017, arXiv: 1611.07725. [Online]. Available: <http://arxiv.org/abs/1611.07725>
- [45] G. Tahir, S. Abrar, and L. Kiong, "A secure scalable life-long learning based on multiagent framework using cloud computing," *Lecture Notes in Networks and Systems*, vol. 70, pp. 400–421, 2020.
- [46] D. Lopez-Paz and M. A. Ranzato, "Gradient Episodic Memory for Continual Learning," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 6467–6476. [Online]. Available: <http://papers.nips.cc/paper/7225-gradient-episodic-memory-for-continual-learning.pdf>
- [47] Y. Xiang, Y. Miao, J. Chen, and Q. Xuan, "Efficient Incremental Learning Using Dynamic Correction Vector," *IEEE Access*, vol. 8, pp. 23 090–23 099, 2020.
- [48] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *arXiv:1612.00796 [cs, stat]*, Jan. 2017, arXiv: 1612.00796. [Online]. Available: <http://arxiv.org/abs/1612.00796>
- [49] Z. Li and D. Hoiem, "Learning without Forgetting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [50] D. Maltoni and V. Lomonaco, "Continual learning in single-incremental-task scenarios," *Neural Networks*, vol. 116, pp. 56–73, Aug. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0893608019300838>
- [51] P. Ruvolo and E. Eaton, "ELLA: An Efficient Lifelong Learning Algorithm," in *International Conference on Machine Learning*. PMLR, Feb. 2013, pp. 507–515, iSSN: 1938-7228. [Online]. Available: <http://proceedings.mlr.press/v28/ruvolo13.html>
- [52] H. B. Ammar, E. Eaton, P. Ruvolo, and M. E. Taylor, "Online multi-task learning for policy gradient methods," in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ser. ICML'14. Beijing, China: JMLR.org, Jun. 2014, pp. II–1206–II–1214.
- [53] C. Clingerman and E. Eaton, "Lifelong Learning with Gaussian Processes," in *Machine Learning and Knowledge Discovery in Databases*, M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, Eds. Cham: Springer International Publishing, 2017, vol. 10535, pp. 690–704, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-71246-8_42
- [54] S. Lee, J. Stokes, and E. Eaton, "Learning Shared Knowledge for Deep Lifelong Learning using Deconvolutional Networks," pp. 2837–2844, 2019. [Online]. Available: <https://www.ijcai.org/Proceedings/2019/393>
- [55] S. Lee, S. Behpour, and E. Eaton, "Sharing Less is More: Lifelong Learning in Deep Networks with Selective Layer Transfer," Jun. 2020. [Online]. Available: <https://openreview.net/forum?id=G8NSCGPEJM>
- [56] T. Zhang, G. Su, C. Qing, X. Xu, B. Cai, and X. Xing, "Hierarchical Lifelong Learning by Sharing Representations and Integrating Hypothesis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–11, 2019, conference Name: IEEE Transactions on Systems, Man, and Cybernetics: Systems.
- [57] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006. [Online]. Available: <https://doi.org/10.1162/neco.2006.18.7.1527>
- [58] H.-E. Kim, S. Kim, and J. Lee, "Keep and learn: Continual learning by constraining the latent space for knowledge preservation in neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11070 LNCS, pp. 520–528, 2018.
- [59] A. Patra and J. A. Noble, "Incremental Learning of Fetal Heart Anatomies Using Interpretable Saliency Maps," in *Medical Image Understanding and Analysis*, ser. Communications in Computer and Information Science, Y. Zheng, B. M. Williams, and K. Chen, Eds. Cham: Springer International Publishing, 2020, pp. 129–141.
- [60] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "CheXpert: A large

chest radiograph dataset with uncertainty labels and expert comparison,” in *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*. AAAI Press, 2019, pp. 590–597. [Online]. Available: <https://jhu.pure.elsevier.com/en/publications/chexpert-a-large-chest-radiograph-dataset-with-uncertainty-labels>

- [61] Z. Li, “lizhitwo/LearningWithoutForgetting.” Apr. 2021, original-date: 2016-10-26T22:04:14Z. [Online]. Available: <https://github.com/lizhitwo/LearningWithoutForgetting>