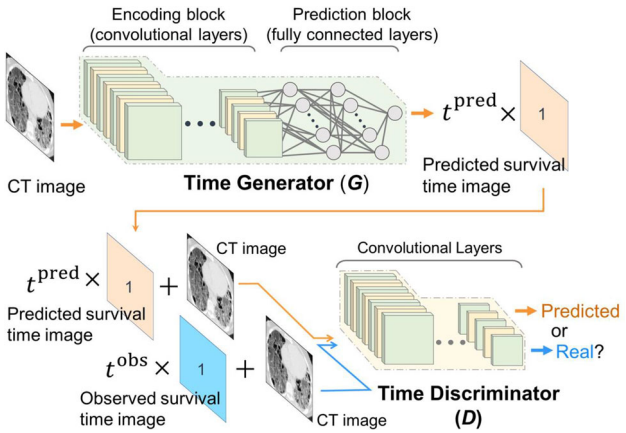

Fig. 1 The schematic architecture of our weakly unsupervised cGAN-based survival prediction model, pix2surv

"observed pairs" of the input CT images and their observed (actual) survival-time images of the patients. The training of vox2surv involves the optimization of G and D through a modified min–max objective function so that G can learn to generate a survival time that is close to the observed survival time. The predicted survival time of a patient is calculated as the median of the predicted survival times of the patient's CT images.

We used the concordance index (C-index) as the metric to evaluate the performance of the prognostic prediction. A bootstrap method with 50 replications was used to estimate the C-index. The prognostic prediction performance of pix2surv was compared with those of the GAP index and the CPI using a two-sided unpaired t-test. We also compared the equivalence of Kaplan–Meier (KM) survival curves generated by pix2surv, GAP, and CPI to the actual survival curve of the patient cohort by using a non-parametric equivalence test.

Results

Table 1 shows the median C-index values calculated from the bootstrap evaluation, which shows that the performance of the mortality prediction by the pix2surv model (C-index: 82.6% [95% CI: 80.8, 83.8]) was statistically significantly higher (bootstrap t-test, $p < 0.0001$) than those of the CPI (63.0% [61.7, 64.3]) and the GAP index (65.9% [64.7, 67.1]).

Figure 2 shows the KM survival curves predicted by the CPI, GAP index, and pix2surv for the mortality of the IPF patients in magenta, red, and green colors, respectively, where, in comparison, the actual survival curve of the patient cohort is shown in blue. Non-parametric equivalence test showed that the survival curves predicted by pix2surv were statistically equivalent to the actual survival curves over the period of 0 to 5000 days, whereas those predicted by the CPI and GAP were not. Also, visual assessment indicates that pix2surv approximates the actual survival curve substantially better than do CPI and GAP.

Table 1 The C-index values estimated by the bootstrap evaluation of the CPI, GAP, and vox2pred. 95% CI = 95% bootstrap confidence interval. *Two-tailed t-test

Clinical/Radiomic Features	C-index (%) [95% CI]	P-value*
composite physiologic index (CPI)	63.0 [61.7, 64.3]	P < 0.0001
Gender, age, and physiology (GAP) index	65.9 [64.7, 67.1]	P < 0.0001
pix2surv	82.6 [80.8, 83.8]	

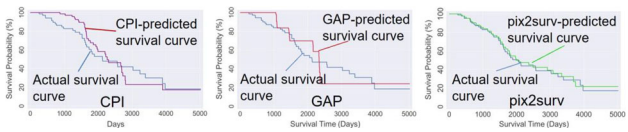

Fig. 2 Predicted KM survival curves by the CPI, GAP index, and pix2surv in comparison with the actual survival curves for the patient cohort

Conclusion

We evaluated the performance of our image-based weakly supervised survival prediction model, pix2surv, which can directly predict the survival of patients from their chest CT images, on the survival prediction of patients with PF-ILD. We showed that pix2surv outperforms the current standards of GAP and CPI in predicting the survival of patients with PF-ILD, indicating that pix2surv can be an effective prognostic biomarker for PF-ILD.

Acknowledgments

This study was supported partly by the NIH/NIBIB/NCI grants of R01HL164697 (PI: Yoshida).

References

[1] Fischer A, Antoniou KM, Brown KK, Cadranel J, Corte TJ, Du Bois RM, Lee JS, Leslie KO, Lynch DA, Matteson EL, Mosca M, Noth I, Richeldi L, Strek ME, Swigris JJ, Wells AU, West SG, Collard HR, Cottin V (2015) An official European Respiratory Society/American Thoracic Society research statement: Interstitial pneumonia with autoimmune features. European Respiratory Journal 46:976–987.

[2] Uemura T, Näppi JJ, Watari C, Hironaka T, Kamiya T, Yoshida H (2021) Weakly unsupervised conditional generative adversarial network for image-based prognostic prediction for COVID-19 patients based on chest CT. Medical Image Analysis 73:1–14.

Application of Convolutional Neural Network with Transfer learning to pattern recognition of tuberculosis in chest X-ray images

L. Lins de Lima[1,2], R. Rodrigues[1], J. Saito[3], O. Yussuf[1], L. Peron[1], K. Lopes[1], M. Koenigkam-Santos[1,4], P. Azevedo-Marques[1]

[1]University of Sao Paulo—USP, Ribeirao Preto Medical School—Medical Imaging, Hematology and Oncology, Ribeirao Preto, Brazil [2]University of Sao Paulo, Interunit graduate program in Bioengineering, Sao Carlos, Brazil [3]University of Sao Paulo, Medical Physics, Ribeirao Preto, Brazil [4]University of Sao Paulo, Bauru Medical Course, Bauru, Brazil

Keywords Tuberculosis, X-ray images, Convolutional Neural Network, Transfer Learning.

Purpose

Tuberculosis is the disease responsible for most deaths provoked by an infectious agent, and it caused around 2.5% of deaths worldwide in 2004 [1]. Tuberculosis can also increase the risk of lung malignancy, chronic pulmonary aspergillosis, and septic shock, even in patients that survived the primary infection caused by the "Mycobacterium tuberculosis," responsible for this illness [1]. The treatment is successful at about 85%, and the mortality rate is about 15% [1].

In diagnosing pulmonary diseases, including tuberculosis, the initially requested imaging exam is the chest X-ray. Although it is considered a simple exam, its evaluation can sometimes be complex. To support radiologists" decision-making, computer-aided diagnosis (CAD) systems have been developed to act as a second opinion through a computer-supplied suggestion.

The use of artificial intelligence (AI) to support diagnostic decision-making in radiology has grown exponentially in recent years. Machine learning (ML) has been the basis for CAD systems. ML is a way of "training" an algorithm so that it can learn. "Training" involves providing large amounts of data to the algorithm, allowing it to adjust and improve its performance. ML algorithms look for patterns within a dataset.

From 2010 to 2012, an approach called Deep Learning (DL) was increasingly adopted for solving ML problems. One approach based on DL is the Convolutional Neural Network (CNN), focusing on classifying images. Modeling the best CNN architecture for any situation by hand can be exhausting, time-consuming, and expensive. Another difficulty is getting a database with a large number of images. However, an alternative option is the Transfer Learning technique, which uses a network pre-trained on a large dataset, for example, the ImageNet [2].

This work aimed to develop a CNN model using transfer learning to support the diagnosis of tuberculosis in chest radiographic images in frontal projection.

Methods

Our institutional review board approved this retrospective study with a waiver of the patient's informed consent (CAAE: 25,762,319.7.0000.5440). For the application of CNN models, a database of chest X-ray images was used with a total of 547 images, with 382 images presenting normal cases and 165 images from patients diagnosed with tuberculosis at the Ribeirao Preto Clinics Hospital. The database was structured with radiographic exams saved in.png format and with three channels (RGB).

The processing environment consists of a server that was accessed remotely; this server has an NVIDIA® TeslaTM T4 16 GB GPU and operates with the Linux 18.04 LTS system. For the execution of the codes, the Jupyter Notebook software was used, which is an interface to browse the server files and execute the codes in the Python language (version 3.9) using the TensorFlow framework (version 2.8).

Five types of transfer models based on CNN were tested. The selected architectures were: VGG19, InceptionV3, DesNet201, ResNet125v2, and XceptionV3. For the use of the models, a global average pooling layer was added to the output of the last layer in each architecture. The original fully connected layers were removed and replaced by only one fully connected layer with two output neurons using the activation function Softmax. All the weights of the models were retrained. These architectures were trained individually, and a standard preprocessing approach with the function "preprocess_input" was performed on the original images to evaluate each model. The number of epochs used was 100. The callback function ReduceLROnPlateau was applied to find the best learning rate value on the validation dataset during the training with values of 0.5, 5, 'max,' and 0.00001 for factor, patience, mode, and min_lr, respectively. The batch size used was 30, and the optimizer function used was Adam to minimize the categorical cross-entropy.

To evaluate the models' performance, the samples were shuffled. The Holdout method was applied to train the networks and classify the images, where 481 images were used for training, 10 percent were separated for validation, and 66 were used for testing the architectures. The models' performance was evaluated based on specificity, sensitivity, and accuracy.

Table 1 Results obtained for each tested model

Model	Specificity (%)	Sensitivity (%)	Accuracy (%)
VGG19	90	90	90
Inception V3	100	81	90
Desnet201	96	81	89
Xception	57	96	77
Resnetl25v2	33	96	65

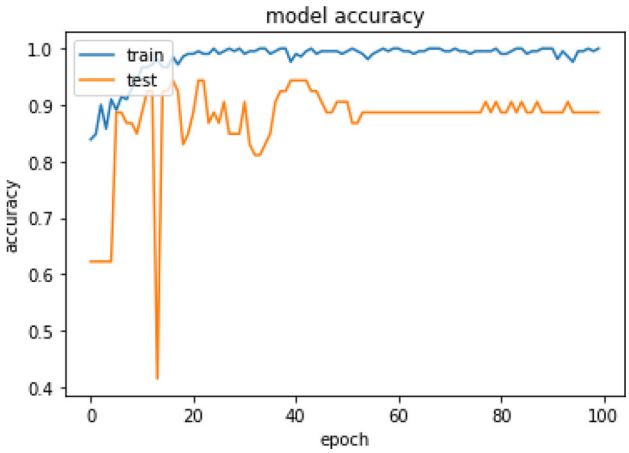

Fig. 1 Accuracy values obtained during the training and testing phases of the VGG19 architecture

Results

At the end of each architecture's training and testing, results presented in Table 1 were generated. During the training and testing phases, the VGG19 model gave the best and most stable results, with specificity, sensitivity, and accuracy of 90%.

Figure 1 presents the accuracy values during the training and testing phases of the VGG19 model.

Conclusion

Based on the results obtained, it can be stated that the networks performed very well when classifying positive images for tuberculosis. It can be seen that the Resnet125v2 and Xception networks had an excellent performance when evaluating the normal cases; however, when considering the images with a positive result for tuberculosis, the VGG19, InceptionV3, and Desnet201 networks presented better results.

In general, the networks presented a satisfactory performance for classifying tasks, and other tests will be carried out using public databases to simulate situations of exams from different sources to verify their generalization power.

Acknowledgments

This study was financed in part by the Coordenacção de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Finance Code 001 and by The São Paulo Research Foundation (FAPESP) —Grants #2020/07200-9, #16/17078-0 and 14/50889-7.

References

[1] Alzayer Z, Al Nasser Y. (2022). Primary Lung Tuberculosis. Treasure Island (FL): StatPearls Publishing. PMID: 33,620,814.

[2] Lima L L, Koenigkam-Santos M, Peron L, Fontes Lopes K L, Lourenço Repolês, M, Koenigkam-Santos M, Azevedo-Marques, P. (2022). Classifying pulmonary interstitial opacities in chest X-ray images using convolutional neural network and transfer learning. In: CARS 2022—Computer Assisted Radiology and Surgery Proceedings of the 36th International Congress and Exhibition Tokyo, Japan, June 7–11, 2022. Int J CARS (2022) 17 (Suppl 1): S93-S94. https://doi.org/10.1007/s11548-022-02635-x

Distal radius fracture classification on dual-view radiography using ensemble deep learning framework

H. Min[1,2], A. Wadhawan[3], Y. Rabi[4], P. Bourgeat[1], J. Dowling[1,2,5,6,7], J. White[3,8], A. Tchernegovski[9], B. Formanek[10], M. Schuetz[4,11,12,13], G. Mitchell[3,8,11], F. Williamson[3,8,11], C. Hacking[3,8], K. Tetsworth[3], B. Schmutz[4,11,12,13]

[1]CSIRO, Australian e-Health Research Centre, Herston, Australia [2]University of New South Wales, South Western Clinical School, Sydney, Australia [3]Royal Brisbane and Women's Hospital, Herston, Australia [4]Queensland University of Technology, School of Mechanical, Medical and Process Engineering, Brisbane, Australia [5]University of Wollongong, Centre for Medical Radiation Physics, Wollongong, Australia [6]The University of Sydney, Institute of Medical Physics, Sydney, Australia [7]University of Newcastle, School of Mathematical and Physical Sciences, Newcastle, Australia [8]University of Queensland, Medical School, Brisbane, Australia [9]Monash Medical Centre, Clayton, Australia [10]University of Queensland, St Lucia, Australia [11]Jamieson Trauma Institute, Brisbane, Australia [12]Queensland University of Technology, ARC Training Centre for Multiscale 3D Imaging, Modelling, and Manufacturing, Brisbane, Australia [13]Queensland University of Technology, Centre of Biomedical Technologies, Brisbane, Australia

Keywords Distal radius fracture, X-ray, Deep learning, Ensemble learning.

Purpose

Distal radius fractures (DRFs) are one of the most common fractures treated surgically. To examine the injured wrists, standard radiographs, including posteroanterior (PA) and lateral views, are often taken in the emergency department. DRFs can be classified into intra- and extra-articular fractures. In extra-articular fractures, the fracture line does not extend to the joint, while the intra-articular fractures involve the articular surface, which may require further evaluation and more complex treatments. Identifying DRFs as intra- or extra-articular can be useful for guiding further treatment. However, radiographic classification of DRFs is challenging due to the extreme variability of fracture patterns, complex anatomy of the wrist and variability in imaging quality of radiography. The aim of this study is to propose a deep learning (DL) framework incorporating both PA and lateral view X-rays for automatic DRF classification and evaluate the framework on clinically acquired wrist X-ray dataset.

Methods

The proposed framework consists of a distal radius region of interest (ROI) detection stage and a DRF classification stage as shown in Fig. 1. The distal radius ROI detection stage used an ensemble model of 10 YOLOv5 [1] base networks which is a recent release of the YOLO object detection network. This step allows the framework to zoom in on the relevant regions on PA and lateral view X-rays for fracture pattern analysis. Following the ROI extraction, an ensemble model of 10 dual-branch EfficientNet (DB-EffiNet) was applied to classify the DRFs into intra- or extra-articular fracture. The DB-EffiNet is a novel adaptation of the EfficientNet [2] constructed in this study, which consists of two EfficientNet-b0 branches taking PA and lateral view X-rays as input respectively. The two branches were fused at the last linear layer via summation, followed by an additional linear layer to generate the final classification output.

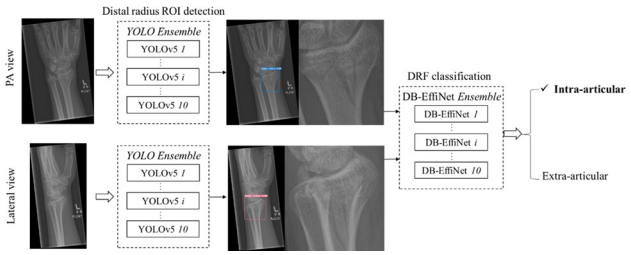

Fig. 1 Diagram of the DL-based DRF classification framework using both PA and lateral view X-rays. DB-EffiNet stands for dual-branch EfficientNet

The dataset used for evaluating the DL framework contains 302 cases of clinically retrieved wrist X-rays. The dataset was randomly split into a training set of 251 cases with 257 fractures and a testing set of 51 cases with 52 fractures. There are 193 and 38 intra-articular DRFs in training and testing set respectively. The training set was randomly partitioned into 10 folds for cross-validation.

For distal radius ROI detection, the YOLOv5s variant was trained on the PA and lateral view X-rays separately for 100 epochs within tenfold cross-validation, generating 10 YOLO base models for each view. The batch size was set as 8 and image size as 1280 × 1280, with stochastic gradient descent (SGD) used as the optimizer. Translation, scaling, horizontal flip and mosaic augmentations were adopted during training. For DRF classification, the DB-EffiNet was trained on the PA-lateral ROI pairs for 50 epochs within tenfold cross-validation. Each EfficientNet-b0 branch was pre-loaded with ImageNet pretrained weights. The ROI images were resized to 256 × 256 and normalized to the mean and standard deviation of ImageNet. The Adam optimizer was used with a learning rate of 0.0001 and the batch size was set as 16. Horizontal flip, rotation and brightness adjustment based on Power–Law transformations were used for augmentation during training. The model with the best area under the receiver operating characteristic curve (AUROC) on the validation set was saved as the base model in each cross-validation iteration.

Given an unseen testing instance, the distal radius ROI was detected on each view through merging the 10 YOLO base models by enabling the model ensemble feature of YOLOv5. The ROIs on the PA and lateral views were then passed into the 10 DB-EffiNet base models. The ensemble probability was computed by averaging the probabilities across all base models.

Results

When evaluated on the testing data, the YOLO ensemble model successfully detected all distal radius ROIs on PA and lateral view X-rays with no false positives. As for differentiating intra- from extra-articular DRFs, the DB-EffiNet ensemble model achieved an AUROC of 0.90, an accuracy of 0.87, a sensitivity of 0.87 and a specificity of 0.86.