CHEST X-RAY CLASSIFICATION USING SELF-SUPERVISED LEARNING

Gowtham B*, Karthikeyan S. **, Pon Harshavardhanan***, Naveen Kurra*, Sumathi D.*[1]
*VIT-AP University-Amaravati, Andhra Pradesh-India,
**KPR Institute of Engineering and Technology, Coimbatore
*** VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore Madhya Pradesh - 466114

ABSTRACT
Deep learning models have created a tremendous impact in a medical image classification problems, especially in the case of chest radiographs. So far, the pathologies in Chest X-Ray images were classified largely using the supervised methodology in which the model learns from both the data and the corresponding labels available. These models require large amount of data in order to perform significantly well and it is very difficult to create large datasets of images with labels particularly medical images like Chest X-Rays. But in recent times, models trained using an unsupervised learning mechanism called self-supervised learning has been performing on par with models trained with the help of supervised learning, where you don't need a large labeled dataset to train a good model. In this paper, shows how models that are trained with self-supervised mechanism on large unlabeled Chest X-Ray images outperform models which are trained with the help of supervised transfer learning in the classification task.

KEYWORDS: Deep Learning, Self Supervised Learning, Image Classification, Unsupervised Learning, Chest X-Ray

MSC: 68T10, 68T30, 68T45

RESUMEN
Los modelos de aprendizaje profundo han creado un tremendo impacto en los problemas de clasificación de imágenes médicas, especialmente en el caso de las radiografías de tórax. Hasta ahora, las patologías en las imágenes de rayos X de tórax se clasificaron en gran medida utilizando la metodología supervisada en la que el modelo aprende tanto de los datos como de las etiquetas correspondientes disponibles. Estos modelos requieren una gran cantidad de datos para funcionar significativamente bien y es muy difícil crear grandes conjuntos de datos de imágenes con etiquetas, en particular imágenes médicas como radiografías de tórax. Pero en los últimos tiempos, los modelos entrenados con un mecanismo de aprendizaje no supervisado llamado aprendizaje auto-supervisado se han desempeñado a la par de los modelos entrenados con la ayuda del aprendizaje supervisado, donde no se necesita un gran conjunto de datos etiquetados para entrenar un buen modelo. En este artículo, se muestra cómo los modelos que se entrenan con un mecanismo auto-supervisado en grandes imágenes de rayos X de tórax sin etiquetar superan a los modelos que se entrenan con la ayuda del aprendizaje de transferencia supervisado en la tarea de clasificación.

PALABRAS CLAVE: Aprendizaje Profundo, Aprendizaje Auto-supervisado, Clasificación De Imágenes, Aprendizaje No Supervisado, Radiografía De Tórax

1. INTRODUCTION

Machine learning algorithms influence most of the decision making in almost all the fields these days. Medical image analysis is one such area where there has been a rapid increase in the adoption of machine learning based disease identification from Chest X-Rays, CT Scans and MRIs over the past decade. This is hugely because of deep learning, which is a subset of machine learning that uses artificial neural networks to learn hidden features and representations from the data. One such artificial neural network called convolutional neural network is very good at capturing patterns in images and are thus extensively used in almost all image classification, segmentation and object recognition tasks. And because of this, different architectures of CNNs are widely used to detect lung and heart diseases from Chest X-Rays [15] and CT scan images [3], segment brain tumors [13] from MRIs, detect and track the growth of cancer cells in skin, etc.
All these achievements over the last few years has been possible because of publicly available large labeled datasets of medical images, increased computation power and larger and deeper CNNs models. This has

[1] Corresponding Author: skarthi.sns@gmail.com

enabled these models to perform better than radiologist and doctors in some of the tasks and also help them in interpretation and decision making.
However, all these models are trained using supervised learning mechanism, where you need both the data of images as well as their corresponding labels. And hundreds of thousand or even millions of images with labels are needed to train these models in order to gain human level performance. Creating a large enough corpus of labeled images is a very tedious task and requires a lot time and effort to annotate images. This issue is profound in the case of medical imaging, where it is very difficult and expensive to create a large dataset of images with labels to train a good model. So, for most of the small medical image datasets containing Chest X-Rays use a supervised pre-trained model that is already trained on millions of images with labels, which in most of the cases we use models that are trained on ImageNet [8] and then fine tune it to a smaller dataset.
This method called transfer learning is pretty useful and causes a considerable increase in the performance of the model. Although this might increase the performance of models on Chest X-Ray images too [10], the Chest X-Ray classification task is different from normal image classification. This is largely because both these domains are different and in the case of Chest X-Rays which are mostly greyscale images, only a small number of pixels account for the presence of abnormalities. Thus, for Chest X-Ray images, there is a need to develop models that doesn't require large labeled datasets and also a different pre-training method to learn useful representations from large unlabeled data which can then be used to train better models with small amount of labeled data.

2. OVERVIEW OF LITERATURE

Ever since deep learning and convolutional neural networks became famous, classifying diseases in chest X-ray images using CNNs has been a very hot topic and a lot many works have been carried out previously. Most of the proposed approaches either use supervised learning or semi supervised learning while only a very few focuses on unsupervised learning methods for classification. Different approaches use various CNN architectures[14] [4], but bigger and deeper models like ResNet and DenseNet have been very popular. Tang el al. [17] analyzed performance of different models like AlexNet, Vgg, ResNet on different datasets which include NIH Chest X Ray[19] dataset [20], Indiana Dataset, WCMC pediatric dataset to predict normal vs abnormal chest X rays. It was found out that the size of the model doesn't impact binary classification in smaller datasets but bigger models were required when the size of the dataset is large. These big models also perform well than others in multiclass classification.
More specifically, transfer learning along with fine tuning using models like DenseNet-121 have increased the accuracy and AUC score significantly in multiclass classification problem [14]. Some other methods use attention mechanism along with CNNs to improve AUC score. Wang et al. [19] proposed ChestNet which has a classification branch that uses CNNs for feature extraction and an attention branch which tries to capture label dependency for abnormality detection. This model was used to detect 14 different thoracic diseases in Chest X-Ray 14 dataset. Another approach used residual attention blocks on the same dataset which used attention to discriminate irrelevant features and maximize correlated feature outputs [6].
Apart from Chest X-Rays, radiologists use several other parameters to detect lung diseases. Taking this into consideration, features like patient age, gender and view position of the X-Ray were used along the images. This integration of non-image features and images seems to have given considerable increase in performance of the model [1]. The ongoing covid-19 pandemic has also pushed researchers to try and detect covid-19 with chest x-rays. For advanced covid testing CT scans are used but since this is costly, chest x-rays are a good alternative which are cheap and CNNs perform very well on chest x-rays than CT scans. XCONVNet [12], a simple CNN model with just 4 convolution layers was able to achieve an accuracy and F1 score of 97 percent. COVID-SDNet [16], another model that used transfer learning on ResNet50 achieved an accuracy of 97.72% and 61.80% in severe and mild COVID-19 cases. DeepChest [7], a multiclass classification model, uses a combination of CT scan and chest x-ray images to detect covid-19, pneumonia and lung cancer at early stages with an accuracy of 98%.

3. METHODOLOGY

In this paper, we use the contrastive learning based SimCLR [18] approach for the classification of pneumonia vs normal chest x-ray images. In general, our approach is divided into two parts. Firstly, we perform the pretext task, where a model learns useful representations from the unlabeled chest x-ray images.

In this part, we employ the SimCLR based self-supervised learning methodology, where the images are split into different batches and each image in a batch is augmented with positive and negative pairs and the model learns to distinguish the positive and negative pairs. For this purpose, we used the NIH Chest X-Ray dataset [19] which has around 112,120 X-ray images from 30,805 unique patients. This is explained in detail in section 4.1. Secondly, once the model has been trained well for a long time during the pretext task, the base encoder network along with its learnt representations are used in the downstream task which is to classify normal vs pneumonia patients from the chest x-ray images using supervision which is explained in section 4.2. Also, we performed this experiment using three different CNN architectures which are ResNet50 and ResNet101, and evaluated the performance of each of these on the final downstream task. For the downstream task, we have used chest x-ray dataset from cell [11] and NIH Chest X-Ray [19]. An overview of the proposed methodology is illustrated in figure 1.

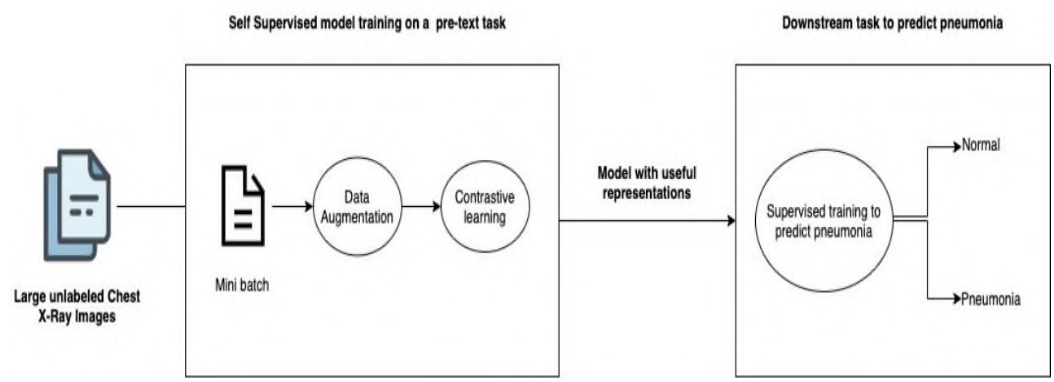

Fig 1. Architecture diagram of the proposed experiment

4. EXPERIMENT

In this section, we describe in detail about model training during the pretext task, parameters involved and the loss function used. We then go on to explain about how this model is used in the downstream task where we do supervised classification.

SELF-SUPERVISED TRAINING: The main aim of self-supervised training using a pretext task is to make the model learn useful representations and features from the unlabeled data. This is done with the help of contrastive learning which tries to differentiate between similar and dissimilar images. For this we use the SimCLR framework which is simple and doesn't require any specialized architectures like other methods. It consists of three parts, data augmentation module, a projection head and a contrastive loss function. We use the unlabeled images from the NIH chest x-ray dataset for this purpose since it is large and has different pathologies which can be used to learn good representations by the model. As mentioned already, we try this with three different CNNs as base feature encoders namely ResNet50 and ResNet101. Firstly, the dataset is divided into different batches with a batch size of N=64. Then an image in a batch is taken and random transformations are applied to it to a get a pair of two augmented images xi and xj. The transformations we used were crop and resize, color jitter, flipping and gaussian blur. We used these transformations because as per SimCLR findings, these are shown to give better accuracies. Then these augmented pairs are passed on to a base encoder network f(x) which is a ResNet.

The encoders have same architecture and shared weights which try to learn representations from the image pairs and produce embedding vectors hi and hj. These are then passed on to non-linear projections heads which are nothing but two fully connected layers with a relu activation function. This is said to work well with the contrastive loss and improves the quality of representations learned. They project the embedding vectors hi and hj into latent space representations zi and zj. Finally, the contrastive loss is applied to these two vectors which uses the cosine similarity to find the similarity between these two vectors. We use NT-Xent (the normalized temperature-scaled cross entropy loss) loss [9] for this purpose which is defined below

$$l_{i,j} = -\log \frac{exp\ (sim(z_i\ ,z_j)/\tau)}{\sum_{k=1}^{2n}\ \ l_{k\neq i}\ exp\ (sim(z_i,z_k)/\tau)} \quad (1)$$

Here n is the batch size, sim(zi, zj) and sim(zi, zk) are the cosine similarities and τ is the temperature parameter. With the help of this loss function and backpropagation, the network learns useful representations to increase the similarity between positive pairs and make the negative pairs distant. For this experiment, τ = 0.1 gave us better accuracies in the downstream task for all the CNN architectures used and we have trained the models for 400 epochs. After the training process, we use the base encoder networks for our downstream task. The entire pretext task is summarized in figure 2.

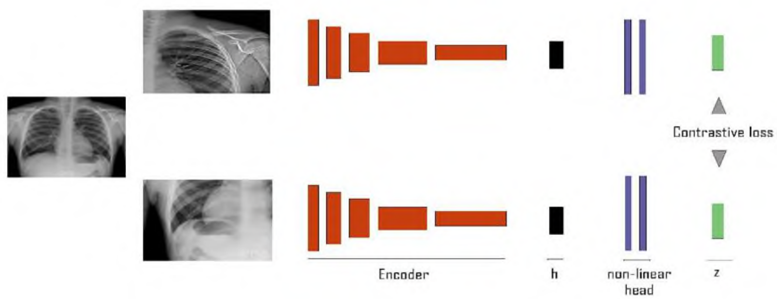

Fig 2. Self-supervised model training on pretext task using contrastive loss

DOWNSTREAM CLASSIFICATION TASK: The base encoder network trained on the above-mentioned pretext task has learned useful representations from the unlabeled chest x-ray dataset. This is now used in the supervised downstream task, where we classify normal vs pneumonia images. For this we use chest x-ray dataset from cell which contains 5863 images of containing normal and pneumonia patients and images of patients having pneumonia and no findings from NIH Chest X-Ray. To this base network, we add two fully connected layers along with a dropout on each and a final sigmoid layer for binary classification. The model is then trained for 30 epochs with a batch size of 16 and learning rate of 0.0001 with Adam as optimizer. Apart from this, we also train 3 other models with the same architectures using supervised transfer learning method without any self-supervised training. Finally, we compare these two methods by calculating the accuracy and the AUC score of all the models.

5. RESULTS

Self-supervised models like SimCLR [18] , BYOL [2] and DINO [5] have been able to perform equally well compared to supervised ones and increasing the width of the models, batch size and number of epochs also seems to give better accuracies on classification task than supervised models on ImageNet data. Our experiments on chest x-ray images show the same as models trained using self-supervised methodology performed better than supervised transfer learning models. We have used two datasets to evaluate the performance and self-supervised models performed well than supervised ones on both the datasets. The models ResNet50 and ResNet101 trained using supervised mechanism produced an accuracy of 82% and 86% on cell dataset and for the NIH dataset it was 85% and 90%. But, the models trained using self-supervised learning produced even better accuracy on both datasets with 91% and 94% on cell and 86% and 92% on NIH using the same ResNet50 and ResNet101 architectures.

6. CONCLUSION AND DISCUSSION

Supervised learning, which was pioneering deep learning research in the past decade has now started to show of its limitations and cannot work without massive amounts of data. Because of this, we are moving now towards fully unsupervised learning which has now started to make huge progress due to techniques like self-supervised learning. Also, we humans tend to learn much better with just little knowledge about objects around us and are very good at generalizing this knowledge to new experiences. This also one reason why supervised learning is not good moving forward. Now in the case of medical image diagnosis, it's very expensive to generate images of patients and manually annotate them. Transfer learning and fine-tuning from other datasets ImageNet is also not the best of choices since the learned weights may come from completely different class of images. Self-supervised learning overcomes all these problems and it could pave the way for more self-supervised models across all different medical image classification, segmentation and recognition challenges which could make development and deployment easier too. Still, there are few issues like high computation requirements, lack of consensus among data augmentation methods reduce the usage of self-supervised training. But active research in this area could bring out better self-supervised frameworks that could solve these issues. In our future works, we aim to use other frameworks like MoCo and BYOL and try to implement these for multilabel classification problems in chest x-ray images and also make a comparison of each of the frameworks used.

RECEIVED: AUGUST, 2022.
REVISED: MARCH , 2023.

REFERENCES

[1] BALTRUSCHAT, IVO M., et al. (2019): Comparison of deep learning approaches for multi-label chest X-ray classification, **Scientific reports,** 9.1, 1-10.

[2] CARON, MATHILDE, et al. (2021): Emerging properties in self-supervised vision transformers, **arXiv preprint**, arXiv: 2104.14294.

[3] CHRISTE A, PETERS AA, DRAKOPOULOS D, et al. (2019): Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images, **Invest Radiol**, 54 (10).

[4] DENG, J., DONG, W., SOCHER, R., LI, L., KAI LI et al. (2009): ImageNet: A large-scale hierarchical image database, **IEEE Conference on Computer Vision and Pattern Recognition**, 248-255.

[5] ESTEVA, A., CHOU, K., YEUNG, S. et al. (2021): Deep learning-enabled medical computer vision, **npj Digit. Med**. 4, 5.

[6] GRILL, JEAN-BASTIEN, et al. (2020): Bootstrap your own latent: A new approach to self-supervised learning, **arXiv preprint,** arXiv: 2006.07733.

[7] GUAN, QINGJI, & YAPING HUANG (2020): Multi-label chest X-ray image classification via category-wise residual attention learning, **Pattern Recognition Letters 130**, pp.259-266.

[8] IBRAHIM, DINA M., NADA M. ELSHENNAWY, & AMANY M. SARHAN (2021): Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, **Computers in biology and medicine**, 132.

[9] KE, ALEXANDER et al. (2021): CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-Ray interpretation, **Proceedings of the Conference on Health, Inference, and Learning**.

[10] KERMANY, DANIEL S., et al. (2018): Identifying medical diagnoses and treatable diseases by image-based deep learning, **Cell 172.5**, 1122-1131.

[11] MADAAN, VISHU, et al. (2021): XCOVNet: Chest X-ray Image Classification for COVID-19 Early Detection Using Convolutional Neural Networks, **New Generation Computing**, 1-15.

[12] MOHAMMAD HAVAEI, AXEL DAVY, DAVID WARDE-FARLEY et al., Brain tumor segmentation with Deep Neural Networks, **Medical Image Analysis**, Volume 35(1), 18-31.

[13] RAJPURKAR, PRANAV, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, **arXiv preprint** arXiv:1711.05225.

[14] SOHN, K., (2016): Improved deep metric learning with multi-class n-pair loss objective, in **Proceedings of the 30th International Conference on Neural Information Processing Systems**, 1857–1865.

[15] SUBRATO BHARATI, PRAJOY PODDER(2020) et al., Hybrid deep learning for detecting lung diseases from X-ray images, **Informatics in Medicine Unlocked**, 20(1).

[16] TABIK, SIHAM, et al. (2020): COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on Chest X-Ray images, **IEEE Journal of Biomedical and Health Informatics**, 24.12, 3595-3605.

[17] TANG, YU-XING, et al. (2020): Automated abnormality classification of chest radiographs using deep convolutional neural networks, **NPJ digital medicine** 3.1, 1-8.

[18] TING CHEN, SIMON KORNBLITH, MOHAMMAD NOROUZI, & GEOFFREY HINTON (2020): **Proceedings of the 37th International Conference on Machine Learning, PMLR** 119, 1597-1607.

[19] WANG, HONGYU, & YONG XIA (2018): Chestnet: A deep neural network for classification of thoracic diseases on chest radiography, **arXiv preprint**, arXiv:1807.03058.

[20] WANG, XIAOSONG, et al. (2017): Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, **Proceedings of the IEEE conference on computer vision and pattern recognition**.