Covid-19 X-ray Image Classification

Maneesh Kumar Singh
University of Illinois at Urbana-Champaign
mksingh4@illinois.edu

Raman Walwyn-Venugopal
University of Illinois at Urbana-Champaign
rsw2@illinois.edu

Satish Reddy Asi
University of Illinois at Urbana-Champaign
sasi2@illinois.edu

Srikanth Bharadwaz Samudrala
University of Illinois at Urbana-Champaign
sbs7@illinois.edu

ABSTRACT

As part of CS598 Deep Learning for Healthcare course, we have decided to reproduce and improve FLANNEL model[5] for COVID-19 classification using X-ray images.

1. INTRODUCTION

COVID-19 pandemic has ravaged the world on an unprecedented scale. It has caused loss of millions of lives and long lasting damages on surviving patients. X-ray imaging is very important part of diagnosis of COVID-19 and other pneumonia and is often the first-line diagnosis in many cases. Using deep learning for X-ray classification is an ongoing research area. There are some useful model proposed for COVID-19 classification using X-rays. FLANNEL is one such model proposed by Zhi Qiao *et al.* [5]. FLANNEL has shown to accurately detect COVID-19 using X-ray images even when trained with only 100 available COVID-19 x-ray images. We have chosen this paper to strengthen our understanding of deep learning models and improve the current research.

2. LITERATURE SURVEY

2.1 FLANNEL for COVID-19 detection

FLANNEL model [5] is a classification model proposed for detection of COVID-19 from other pneumonia types and normal x-ray images. In this paper, Zhi Qiao *et al.* has shown that with ensemble learning FLANNEL can detect and diagnose COVID-19 from pneumonia x-ray images with high accuracy, even when trained on only 100 available COVID-19 x-ray images.

FLANNEL model introduces two stage classification, where first stage involves using state of arts CNN models to classify dataset into 4 classes: COVID, Pneumonia virus, Pneumonia bacteria, Normal. As train dataset is very small, pre-trained models using ImageNet are utilized.

In stage-2, a neural weight module is used to learn weights for all five predictions in Stage-1. For model training, instead of cross-entropy loss, focal loss is extended to handle multiclass classification.

$$LossFunc = FocalLoss(\hat{y}, y) \quad (1)$$

$$= \sum_{m=1}^{M} -\alpha_m y_m (1 - \hat{y}_m)^{\gamma} \log \hat{y}_m \quad (2)$$

2.2 COVID-19 classification using chest CT

X. Bai and Wang [1] were able to create an AI system that could differentiate COVID-19 and other pneumonia using a chest CT scan. They approached this as a classification problem and used the EfficientNet B4 architecture which was a CNN based network. They were able to achieve results of 96% accuracy, 95% sensitivity , 96% specificity, and an area under receiver operating characteristic curve of 0.95 and an area under the precision recall curve of 0.90. When compared with radiologists on the same test dataset, the AI system performed better. This study concluded that the AI can support radiologists in detection of COVID-19 in Chest CT images.

2.3 Focal loss for dense object detection

Lin T, Goyal P, Girshick R propose Focal loss [3], a modification to the standard cross entropy criterion that focuses weights for loss on hard examples versus well classified examples. This is accomplished by adding a factor $(1-p_t)^{\gamma}$ to the standard cross entropy criterion where setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$). This results in achieving higher accuracy than using the standard cross entropy loss and surpassed speed and accuracy when compared with state of the art two stage detectors; Faster R CNN Variants.

2.4 Towards contactless patient positioning

This paper [2] discussed about patient positioning routine that comprised a novel robust dynamic fusion (RDF) algorithm for accurate 3D patient body modeling. With 'multi-modal' inference capability, RDF can be trained once and used across different applications (without re-training). They have used multiple CNN branches o learn the joint feature representation and a fully-connected parameter regressor module to estimate the 3D mesh parameters.

2.5 Deep learning COVID-19 features on CXR using limited training data sets

The authors of this paper[4], proposed a patch-based convolutional neural network approach with a relatively small number of trainable parameters for Covid-19 diagnosis. The architecture contains first pre-processed data that are fed into a segmentation network [FC-DenseNet] to extract lung areas. From this segmented lung area, classification network is used to classify the corresponding diseases using a patch-by-patch training and inferences [ResNet-18 (pre-trained) and many ResNet-18 models are used for K patches], final decision is made based on the majority voting from previous layers. A Grad-CAM saliency map is calculated to provide an interpretable result. This method has an accuracy of 91.9%, compared to that of 92.4% for COVID-Net.

2.6 COVID19-Net Deep Convolutional Neural Network

This is the first open source network design for COVID-19 detection from CXR images, our final research paper also considers this as its baseline for experiments. This paper considered COVIDx dataset which contains 13,975 CXR images for training and experiments. COVID-Net architecture makes heavy use of a lightweight residual 'projection expansion projection extension' (PPEX) design pattern that contains multiple levels of convolution layers with fully connected layers and a softmax at the end. COVID-Net achieved higher test accuracy than other architectures such as VGG-19 and ResNet-50.

2.7 Noise-robust segmentation of COVID-19 from CT images

This is a CNN model [6] developed to be effective with detection of COVID-19 lesions from CT images that have a lot of noise. This paper discusses how Wang et al developed a novel noise-robust learning framework based on self-ensembling of CNNs. To better deal with the complex lesions, a novel COVID-19 Pneumonia Lesion segmentation network (COPLE-Net) was proposed that uses a combination of max-pooling and average pooling to reduce information loss during downsampling, and employs bridge layers to alleviate the semantic gap between features in the encoder and decoder. Experimental results with CT images of 558 COVID-19 patients showed the effectiveness of the noise-robust Dice loss function, COPLE-Net and adaptive self-ensembling in learning from noisy labels for COVID-19 pneumonia lesion segmentation. To make the training process robust against noisy labels, a novel noise-robust Dice loss function was proposed and integrated into a self-ensembling framework, where an adaptive teacher and an adaptive student are introduced to further improve the performance in dealing with noisy labels.

3. DATA

For the purpose of reproduction and show comparable improvement, we have decided to use the same datasets that are used in the FLANNEL paper.

1. Covid Chest X-ray (CCX) dataset: This dataset contains COVID-19 pneumonia images as well few X-ray images from other classes. The dataset can be obtained from GitHub.

2. Kaggle Chest X-ray (KCX) dataset: This dataset contains normal, bacterial pneumonia, and nov-COVID-19 viral pneumonia. The dataset can be obtained from Kaggle.

In the FLANNEL paper, 5508 chest x-ray images across 2874 independent patient cases. Both dataset contains anteroposterior (AP) and posteroanterior (PA) view.
As done in the research paper, we will include both AP and PA views. Due to AP and PA views being different types of X-ray images, horizontal flips and random noise will be used to convert PA into AP view.

4. APPROACH

We have planned to reproduce the FLANNEL results using the the updated data from original data sources. So our approach consists of two distinct stages.

4.1 Reproduce FLANNEL

4.1.1 Stage-1: Base Learner Training

As done in the original paper, we would use CNN models InceptionV3, Vgg19_bn, ResNeXt101, Resnet152 and Densenet161 as base learners. Due to limited data for training, we will utilize pre-trained models on ImageNet and fine-tune each model for COVID-19 classification.

4.1.2 Stage-2: Ensemble model learning

We would feed the predictions from base learners to FLANNEL neural weight module to learn base learner weights. For learning, we use the Focal loss function modified for multi-class classification.

4.2 Performance Analysis

We would use below measures to calculate the accuracy of Ensemble model for verification.

1. Precision-Recall curve
2. ROC curve
3. Confusion Matrix
4. F1 score comparison accuracy of all methods

4.3 Improvement Discussion

Based on the performance of the novel FLANNEL architecture, our team is motivated to improve the performance further. With inspiration from the approach of Oh Y., and Park S., a primary improvement that we are proposing is to extract the lung contours from the CXR images prior to classification. The motivation behind this is to have the classifier focus on the specific lung regions versus the whole CXR image. In addition, another improvement we propose is to modify each base model in the ensemble to process the segmented CXR image in K patches.
To accomplish this, the steps we will follow are

1. Extract the lung contours from the CXR images using Segment Model.

2. Train k-patch classifiers.
   (a) Start with pre-trained model from Image Net
   (b) Divide extracted lung contours into k-patches
   (c) Run each patch through model to generate classification, prediction is calculated based on majority voting
   (d) Update shared weights for each of the k models

3. Construct improved FLANNEL architecture by using extracted lung contours as input and k-patch classifiers used as the base models.

4. Train Improved FLANNEL architecture
   (a) Input is CXR image

Table 1: Experimental data description

Source		Total	COVID-19	Viral	Bacterial	Normal
Original data	CCX data	554	478	16	42	18
	KCX data	5856	0	1493	2780	1583
View Distribution	AP view	6163	282	1501	2789	1591
	PA view	247	196	8	33	10

AP: anteroposterior; CCX: COVID Chest X-ray; COVID-19: coronavirus disease 2019; KCX: Kaggle Chest X-ray; PA: posteroanterior.

(b) Extract lung contours using Segment Model

(c) Create k-patches of segmented CXR

(d) Each k-patch classifier processes k-patches and produces predictions

(e) Calculate weighted ensemble through neural weighting module

(f) Compute prediction based on k-patch model predictions and weights

(g) Compute focal loss and update neural weighting module weights

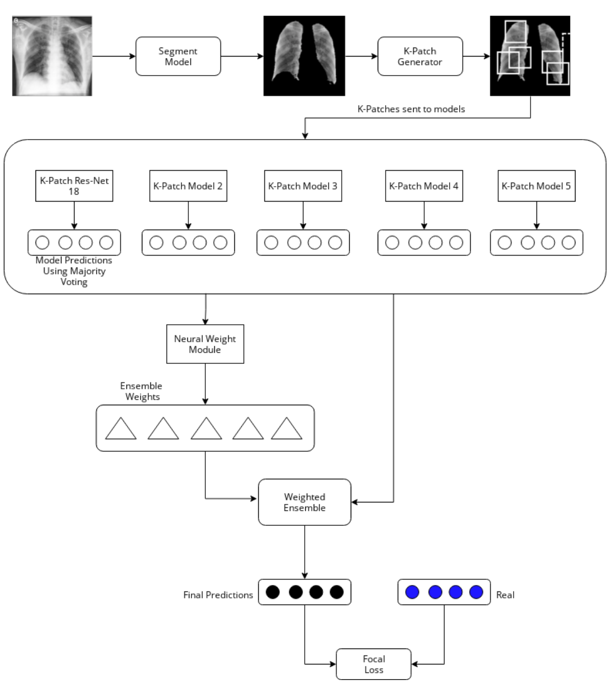
(h) Continuously calculate metrics to measure performance such as but not limited to: accuracy, precision, F1 and ROAUC.

5. Perform ensemble (combining multiple K patch classifiers) to calculate the weighted ensemble.

6. Get the prediction and compare with the ground truth.

7. Apply Focal Loss to train the model (improve weights).

8. Test the model on the 'test' dataset to calculate accuracy, precision, F1 measure, ROAUC and other metrics.



5. EXPERIMENTAL SETUP

We are planning to use FLANNEL source code as our baseline and enhance on top of it. Our codebase would be using below software/python packages

Table 2: Software/Tools used

Software/Tool	Version
Python	3.8.5
numpy	1.20.2
torch	1.8.1
torchvision	0.9.1
matplotlib	3.4.0
scikit-learn	0.24.1
pandas	1.2.3

As FLANNEL model requires significant compute and GPU resources (3 NVIDIA Tesla P100 GPUs). We would be utilizing AWS EC2 service with instance type p3.8xlarge which provides 4 NVIDIA Tesla V100 GPUs along with 32 core CPU and 64GB RAM.

For data analysis and exploration, we are going to use Google Colab.

6. TIMELINE

Table 3: Project Timeline

Task	Planned Due Date
Research, Planning and proposal	03/28/2021
Data Cleaning and Preprocessing	04/02/2021
Extract the lung contours	04/09/2021
Apply classifier on each patch	04/09/2021
Ensemble the results from each classifier	04/16/2021
Calculate weights	04/23/2021
Run the training data to get the prediction	04/23/2021
Apply Focal Loss	04/23/2021
Model Training	04/23/2021
Performance Evaluation	04/23/2021
Documentation and Video presentation	05/07/2021
Code and Report Submission	05/08/2021

7. REFERENCES

[1] H. X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chang, K. Halsey, T. M. L. Tran, J. W. Choi, D. C. Wang, L. B. Shi, J. Mei, X. L. Jiang, I. Pan, Q. H. Zeng, P. F. Hu, Y. H. Li, F. X. Fu, R. Y. Huang, R. Sebro, Q. Z.

Yu, M. K. Atalay, and W. H. Liao. Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT. *Radiology*, 296(3):E156–E165, 09 2020.

[2] S. Karanam, R. Li, F. Yang, W. Hu, T. Chen, and Z. Wu. Towards contactless patient positioning. *IEEE Transactions on Medical Imaging*, 39(8):2701–2710, 2020.

[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection, 2018.

[4] Y. Oh, S. Park, and J. C. Ye. Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. *IEEE Trans Med Imaging*, 39(8):2688–2700, 08 2020.

[5] Z. Qiao, A. Bae, L. M. Glass, C. Xiao, and J. Sun. FLANNEL (Focal Loss bAsed Neural Network EnsembLe) for COVID-19 detection. *Journal of the American Medical Informatics Association*, 28(3):444–452, 10 2020.

[6] L. Wang and A. Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, 2020.