COSC 2673/2793 - Assigment 4

Adeel Ahmed (s3802338)

I. INTODUCTION

COVID-19 has become a global pandemic and is arguably the biggest cause of concern in 2020 [1]. COVID19 is a respiratory disease that is spread from airborne water particles. The disease is said to spread quickly and has affected a large number of people across the world. It has been proposed that chest imaging like CT scans and chest X Rays contain information that can help diagnose COVID-19. The intentions of this paper are to explore this hypothesis, along with exploration of existing work and available datasets, and to provide a solution to try to help the world in these dire times.

II. LITERATURE REVIEW

Our hypothesis states that Chest X Rays contain patterns that are useful for identifying COVID-19 in patients. It is widely discussed that CNNs are extremely useful in pattern recognition, hence our proposed solution will follow said approach. An exploratory analysis was conducted by Yamashita and team [2] that discusses the applications of Convolutional Neural Networks (abbreviated to CNN) in Radiology. Yamashita suggests that CNNs are one of the most dominant methodologies in Computer vision and that it has a large number of applications in radiology. Yamashita discusses CNNs in general along with the architecture and the layers that constitute a network, the author then goes on to discuss other aspects of CNNs, that include (but are not limited to) training, overfitting, applications in radiology, detection, segmentation and so on. Two important takeaways from the paper reside in the Training section along with the challenges section. First the author discusses how CNNs can be made use of even when data is scarce in amounts. This is relevant due to the low availability of COVID-19 related X Ray images. The author suggests that CNNs can first be transfer learnt then made to overfit on the small dataset. Another important takeaway is to note the vulnerability of CNNs to adversarial examples. The author suggests that CNNs tend to give false positives and false negatives when introduced to adversarial examples, although the author goes on to discuss that the impact of adversarial examples in medical conditions is unknown, it is an area which will need attention, this is relevant since the proposed solution is intended to contribute to help during the pandemic, hence the system needs to be evaluated on adversarial examples beforehand, and statements need to be made about it.

A similar research was carried out by Narin and others [3], where CNNs were used to carry out detection of COVID19 cases in patient chest X ray data. The author starts out by discussing the importance of early detection of COVID19 cases and how it can lead to early prevention. The author details this importance with graphics that illustrate the distribution of worldwide cases. The author then goes onto discuss the dataset of choice and the processes being applied to it. The author acknowledges the relatively small size of the dataset with only 100 images. The author proposes the method of deep transfer learning to remedy this, discussing in detail the earlier successes of CNNs when put under similar conditions and restraints. The author describes the experimental setup created for this research, which proposes the use of multiple CNN architectures and their comparisons. The author then goes onto discuss and compare results gathered and then crowns the wining architecture that reportedly produces 98% accuracy. The limitations of this paper are obvious right away. The author used a minimalistic dataset without any augmentation and had a total of 100 images. Thought the author suggests that to avoid overfitting the training was limited to 30 epochs, the training graphs provided indicate that the crowned model had reached an accuracy of 100% quite earlier in the training process, meaning the model had already overfit to the problem. Moreover, the author fails to describe the validation set that was used for this evaluation, hence the robustness of the system will not be the same when put in a real-life scenario.

A paper by Apostolopoulos [4], built further on this hypothesis. The author acknowledged the size of the dataset and started out the paper by describing the process off data collection. The author described how they built two distinct datasets, labelled dataset_1 and dataset_2. It was discussed how pneumonia was also one of the diseases that occur in COVID-19, so it was possible that Chest X Rays of pneumonia affect patients could look similar to Chest X Rays of COVID-19 affected patients. So, for these purposes two distinct datasets (each dataset with or without pneumonia-based X Rays) were created. This detail was very relevant and was the most important takeaway from the paper, this paper provided important domain specific information that any novice researcher may not be knowledgeable of but could impact accuracy of the system. The author then went on two discuss the training of the systems, the author tested 5 different CNN architectures. Each architecture was tested twice (once on each dataset) and different metrics were calculated for each. All the networks were transfer learnt and proved to perform well, with the best model producing up to 98.75% accuracy on the dataset without X Rays of pneumonia patients, and 93.5% accuracy on the dataset with X Rays of pneumonia patients. The author treated the smaller dataset as a binary classification problem and the larger (the dataset that included pneumonia affected Chest X Rays) as a multiclassification problem. The limitations of the paper although minimal, could affect the final outcome of the methodologies discussed. To start with data augmentation could be experimented with to check if improvements in accuracy could be made. Additionally, a thorough evaluation dataset is once again not described, it is important to include chest X rays of not only normal patients, but also those affected by other lung diseases or have lung strenuous habits like smoking etc. The current evaluation is a bit biased since it classifies between healthy and COVID affected X rays as opposed to NON COVID affected and COVID affected X Rays.

III. Proposed Methodology

From the literature review a few things have become evident. To start with our problem is a class imbalance problem. In the reviewed literature it became evident that the number of images for either classes were not equal. Secondly the literature supports our initial proposal that a convolutional neural network is the way to go. Thirdly it could also be seen that the dataset at present is too small to train a model from scratch. Transfer learning is a must and our proposed methodology intends to build upon this but a bit differently. We hypothesize that training a model transfer learnt on image recognition networks may not be as efficient. It could be much more beneficial if a model was trained transfer learnt on medical imaging trained networks. The proposed methodology will be expanded on later, for now a detailed discussion on other aspects will be carried out.

A. Dataset

The dataset [5] that is available is relatively small sized and is of Chest X Rays. The dataset was provided by Victoria Public Hospital and is consists of 200 images of X Rays of COVID-19 patients, along with up to 300 images of healthy patients, and of those who have been affected by other lung diseases (not COVID-19). The images are monochromatic in nature, this can be observed in fig 3.1.

(a) (b) (c)

Fig 3.1 A series of images from the available dataset. Fig (a) shows a direction unaware X Ray. Fig (b) shows a different colored side pose X Ray. Fig (c) shows a direction aware X Ray

The images seem to be monochromatic, but the major color seems to differ. Moreover, the images are variating in size. The pose for the X Ray also differs. Specific preprocessing needs to be carried out before this dataset can be used to train a convolutional neural network.

B. Data Preprocessing

The X Rays were taken out by different doctors, each of which had different machines, this led to high variations in the dataset. This dataset needs to be standardized and normalized before training can begin. The dataset will undergo a series of operations, they are as follows.

a) Color Normalization: All the images are monochromatic [6] meaning that the images are created using different variations of the same hue. A single hue base is extended upon by using various tints, shades, and tones of the same hue. Due to this variation in the dataset color normalization is necesarry. All images will be converted to grayscale. Doing so would put all the images under the same monochromatic color range (the derivative hue would be gray across all images). Not only will this standardize the input but this would also transform the input from multi channel to a single channel, thus decreasing our input data size, which will in turn affect the input layer size, decreasing it.

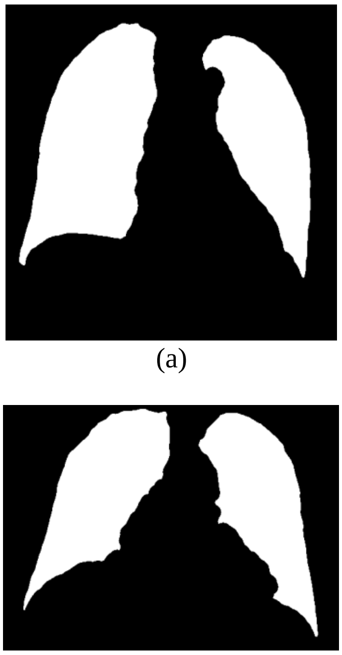
b) Resizing with aspect ratio: The images need to be resized to a common size before they can be made use of. This step is necesary since the input layer of a neural network expects inputs of the same size. To achieve this uniformity many steps can be applied, but each has their pros and cons. Resizing images is a possibility but this tends to skew/stretch the image, since not all images will have the same aspect ratio. This skew or stretch could be major or minor depending on the original resolution of the image and hence is not desirable. Cropping of the image is another possibility but since the cropping will be done with focus on the center there is a possibility that useful sections of the image may get cut out, which would be undesirable. So our best bet is to use resizing but by maintaining the aspect ratio. This can be done by padding the image. This is done be resizing the image on the larger axis (width or height). The image now meets one of the requirements of size but lacks size on the other axis, for these purposes the image is padded (equally distributed across the center) with black pixels. This padding helps to resize the image to a uniform input size while aspect ratio is maintained.

c) Image Normalization: The image pixels need to be normalized to ensure that the pixels of the images have similar data distribution. This is essential when training neural network since it helps the model to converge faster [7]. The image will be normalized so that the pixel values lie between the range 0-1. Images will first be converted to type float and then 255 (highest pixel value) will be added to each pixel in the image, and then all values will be divided by 255. This will range normalize the pixels between 0-1. Standard deviation normalization could also have been carried out which would distributre the pixels over a gaussian curve, but due to the monochromatic nature of our dataset this wont be necesarry.

d) Data Augmentation: The section of preprocessing will be an experimental process to check if the model can attain improved results with augmentation. The major augmentation will be Left Right Flipping, where the negative and positive axis of the image will be swapped, effectively flipping the image in the horizontal direction. This specific augmentation can be done only for the side pose images, this will somewhat increase the number of images in the dataset. Optionally the direction unaware images can also be flipped, but due to the anatomy of the human body it is unsure if this will affect the accuracy of the system. Hence, experiments will be carried out to see if this turns out to be useful or not.

C. Feature Extraction

The X Rays carry some extra information like the arms and/or spinal cord of the patient. This information might not be useful since the disease affects the lungs but may introduce complexity for the CNN when trying to learn about the disease. For these purposes a feature extraction process that is the masking of unwanted areas will be carried out. The dataset carries additional images that correspond to the actual dataset. These images contain a mask of the lungs and can be visualized in Fig 3.2



(a)

(b)

Fig 3.2 Shows different sorts of masks that can be applied to lung images

This mask can be applied to the original image before pre-processing. The mask pixel values will be divided by 255. This will result in only pixel values of 0 or 1. Then a bit-wise multiplication will be done with the original image. The black sections of the mask will null out the same sections of the original image, and the white sections will lead to those corresponding regions being unaffected. Via this process we will get blacked out areas in the resulting image that are not a part of the lung. This masked out image will be the features that will be then used to train the network. This feature extraction will also be experimental and will undergo comparisons with other models to evaluate if this method of feature extraction was useful or not.

D. Proposed Model

A Convolutional Neural Network will be the model of choice for this problem. As discussed earlier it is proven that CNNs are better at pattern recognition when compared to traditional models and is also the top choice for majority of the reviewed literature.

a) Transfer learning: Deep Neural Networks/ CNNs are known to be data hungry machine learning models, but for this problem data is scarce, hence we will be making use of transfer learning. Transfer learning is making use of pre trained models to fit them to your small set of data. If we were to start model training from scratch we would need to wait for the model to first learn feature extraction like shape extraction, texture extraction, etc. After this process the model would then start to solve the classification problem. Not only does this require more data, this also requires more time. By using pre trained models we skip the need for the model to learn feature extraction since we utilize previously trained models. This results in our model training much faster and without the need of large amounts of data.

b) Experimental Setup: A pre-trained model stack of some of the most famous convolutional neural networks will be created. To each model an appropriate classificiation head will be attached. Three independent models will be trained and compared to evaluate the performance and a "best model" will be crowned. This experimental setup can be viewed in figure 3.3.

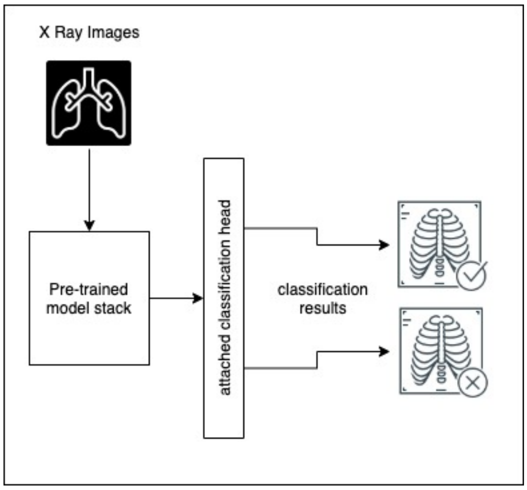


Fig 3.3 Shows the experimental setup that will be prepared

The model stack will consist of namely three models, VGG16 [8], InceptionV4 [9] and Resnet50 [10]. The model specifications can be seen in table 3.1

Model #	**Model Stack**		
	Model Name	***Trainable params***	***Pretrained***
1	VGG16	138 million	Yes
2	Inception V4	43 million	Yes
3	ResNet 50	23 million	Yes

Table 3.1 Shows the model stack of the experimental setup that will be used

c) Evaluation: The original train set will be partitioned, where 80% of the data will be used for the training of the models and the remaining 20% will be used to evaluate the models and to compare them. It will also be made sure that the split is done keeping the class distribution in mind, hence the evaluation set will not be deprived of data from specific classes due to the unbalanced nature of the dataset.

Three different metrics of performance will be used to evaluate the models and to compare them. Namely, accuracy, recall and F1 score.

These performance metrics are required for a sound and unbiased judgement. What we have is an imbalanced class problem, for these reasons specifically accuracy and recall alone are not enough. Hence, we use F1 score as the primary performance metric. F1 score keeps in mind the class distribution and calculates score accordingly, unlike accuracy which tends to give a bias look at the results. Despite this, accuracy and recall are also used as performance metrics to evaluate which models could be a better fit if an increment in the data is observed.

d) Technical Difficulties: Due to the large and deep nature of the neural networks it is possible that we may face certain technical difficulties in training the model. Our models are huge in size, the smallest model having 23 million trainable parameters and the largest having up to 138 million. Training such large models could prove to be daunting for certain computers and could hinder progress or stop training altogether. There's a possibility that the models take too long to load, or cannot be trained on smaller machines altogether. These problems will be resolved by either decreasing the batch size to train the model in smaller incremental steps. This will allow us to train the model but will increase the train time considerably. If these technical difficulties are not resolved then we will resolve to using smaller deep neural network models like Mobile net. MobileNet V2 contains a minimal of 3.47 million parameters and is known to give similar performance (albeit slightly degraded) when compared to full fledged neural networks like VGG etc. Another technical difficulty that could be faced would be out of memory issues, for these purposes we would resort to on the fly data reading and processing as opposed to loading the entire dataset into memory. Additionally data can be processed and stored beforehand at a one time cost to reduce the time complexity of processing.

e) Considerations and Biases: It should be noted that the model is not one without biases, the model could overfit to the type of data or to the nature of data. Meaning when exposed to scans of lesser or even higher quality the performance of the model may output unpredictable results. Additionally it should be noted that the selected machine learning models rely on patterns or shapes to make an educated guess. The model may be sensitive to occlusions and in their presence could return unpredictable results. Occlusions such as pacemakers that are embedded into patients can cause such unpredictable results and hence usage of this system in such scenarios is advised against. The current dataset is that of previously healthy patients, if X rays of previously affected patients or patients with other lung diseases (Datasets where X ray of smokers is also included) are introduced then results may be unpredictable.

f) Ethical Declaration: It should be noted that this system is not meant to be taken as a decisive authority but merely as an aid to one. No decisions should be taken merely on the outputs of this system since it may not perform well in unfavorable scenarios and may lead to unpredictable results.

References

[1] Roosa, K., et al. "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020." *Infectious Disease Modelling* 5 (2020): 256-263.

[2] Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging* 9.4 (2018): 611-629.

[3] Narin, Ali, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks." *arXiv preprint arXiv: 2003. 10849* (2020).

[4] Apostolopoulos, Ioannis D., and Tzani A. Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." *Physical and Engineering Sciences in Medicine* (2020): 1.

[5] Cohen, Joseph Paul, Paul Morrison, and Lan Dao. "COVID-19 image data collection." *arXiv preprint arXiv :2003. 11597* (2020).

[6] Van Gelderen, Fred. *Understanding X-rays*. Berlin: Springer, 2004.

[7] Finlayson, Graham D., Bernt Schiele, and James L. Crowley. "Comprehensive colour image normalization." *European conference on computer vision*. Springer, Berlin, Heidelberg, 1998.

[8] Carvalho, Tiago, et al. "Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN." *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2017.

[9] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." *Thirty-first AAAI conference on artificial intelligence*. 2017.

[10] Akiba, Takuya, Shuji Suzuki, and Keisuke Fukuda. "Extremely large minibatch sgd: Training resnet-50 on imagenet in 15 minutes." *arXiv preprint arXiv:1711.04325* (2017).